

RICE UNIVERSITY

**Transfer of the Testing Effect: Just How Powerful Is It?**

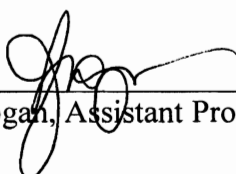
by

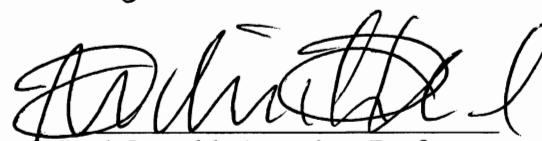
**Gunes Avci**

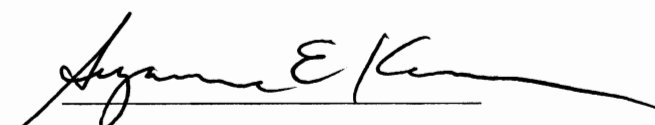
A THESIS SUBMITTED  
IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE

**Doctor of Philosophy**

APPROVED, THESIS COMMITTEE:

  
Jessica Logan, Assistant Professor, Chair  
Psychology

  
Fred Oswald, Associate Professor,  
Psychology

  
Suzanne Kemmer, Associate Professor  
Linguistics

HOUSTON, TX  
APRIL 2011

## ABSTRACT

Transfer of the testing effect: Just how powerful is it?

by

Gunes Avci

Researchers have repeatedly showed that when students take a test on the studied materials, their retention of information increases on later exams. The memory gain produced by tests is called the *testing effect*. The testing effect has been demonstrated with different materials (e.g., word lists, prose), with different designs (e.g., within-subjects, between-subjects), with different age groups and with different settings (e.g., laboratory, classroom). This thesis offers a novel approach to the assessment of the testing effect: the transfer of the testing effect. Although research on the testing effect has been flourishing over the last decade, transfer of the effect has not been studied systematically. In this context, the transfer of the testing effect refers to whether a learner can utilize the memory gain produced by tests in different contexts. The present study examined the transfer of the testing effect in the temporal domain which refers to retention of the testing effect over a period of time (Experiment 1, Experiment 2, Experiment 3 and Experiment 4), the knowledge domain which refers to transfer of the testing effect across different levels of learning (Experiment 1, Experiment 2 and Experiment 3) and the modality domain which concerns how the format of the tests influence the testing effect (Experiment 4). In addition to the laboratory studies, Experiment 3 was conducted in an actual classroom setting incorporating real classroom curriculum. The study revealed that the testing effect endures over a long period of time varying between 1 and 12 weeks and is transferable across different levels of learning (i.e., lower level and higher level). Furthermore, the testing effect is affected by the differences

in the format of tests. The transferability of the testing effect across different domains indicated that the testing effect reflects more than rote learning. The findings of the study were discussed in light of two competing theoretical orientations explaining the testing effect: the retrieval hypothesis and the transfer-appropriate processing framework. Additionally, the practical implications of the study were discussed in relation to the educational system.

# Acknowledgements

First and foremost, I am heartily thankful to my advisor, Dr. Jessica Logan, whose encouragement, supervision, and support at every step of my research enabled me to develop an understanding of the subject. It has been an honor to be her first Ph.D. student. I appreciate all her contributions of time and ideas to make my Ph.D. experience productive and stimulating.

For sharing their time and expertise, I would like to express my appreciation to the members of the committee: Dr. Fred Oswald and Dr. Suzanne Kemmer.

Additionally, I would like to thank the members of the Logan group whose comments, constructive criticisms, and support have contributed immensely to my professional time at Rice University.

Last, but not least, I would like to express my sincere gratitude to my favorite Ukrainian for being moral support and for keeping my spirits high throughout graduate school.

I would like to dedicate my dissertation to my dear family, my father, my mother, and my sister. Without their unconditional support and love, I could not have gotten through this.

# Table of Contents

|   |             |
|---|-------------|
| <b>Acknowledgment.....</b>  | <b>iv</b>   |
| <b>Table of Contents.....</b>   | <b>v</b>    |
| <b>List of Figures.....</b>   | <b>viii</b> |
| <b>List of Tables.....</b>  | <b>ix</b>   |
| <b>Introduction.....</b>  | <b>1</b>    |
| 1.1. The Testing Effect: The Paradigm .....   | 1           |
| 1.2. The Testing Effect in the Laboratory and in the Classroom.....   | 2           |
| 1.3. Theories of the Testing Effect.....  | 7           |
| 1.3.1. Overlearning/Exposure Hypothesis.....  | 7           |
| 1.3.2. Retrieval Hypothesis.....  | 10          |
| 1.3.3. Transfer-Appropriate Processing View.....  | 14          |
| 1.4. Transfer of Learning.....  | 16          |
| 1.4.1. Dimensions of Transfer of Learning.....  | 17          |
| 1.5. Transfer of the Testing Effect.....  | 20          |
| 1.5.1. Transfer in the Temporal Domain.....   | 20          |
| 1.5.2. Transfer in the Modality Domain.....   | 22          |
| 1.5.3. Transfer in the Knowledge Domain.....  | 25          |
| 1.5.4. Evaluation of the theories with respect to the transfer of the testing effect in<br>modality, temporal and knowledge domain..... | 29          |
| <b>The Present Study.....</b>   | <b>30</b>   |
| 2.1.Experiment 1.....   | 35          |
| 2.1.1. Method.....  | 36          |
| 2.1.1.1.Participants.....   | 36          |
| 2.1.1.2.Materials.....  | 36          |
| 2.1.1.3.Design.....   | 37          |
| 2.1.1.4.Procedure.....  | 38          |
| 2.1.2. Results.....   | 39          |

|  |    |
|--|----|
| 2.1.3. Discussion.....                               | 40 |
| 2.2.Experiment 2.....                                | 41 |
| 2.2.1. Method.....                                   | 42 |
| 2.2.1.1.Participants.....                            | 42 |
| 2.2.1.2.Materials.....                               | 43 |
| 2.2.1.3.Design.....                                  | 43 |
| 2.2.1.4.Procedure.....                               | 46 |
| 2.2.2. Results.....                                  | 46 |
| 2.2.2.1.Initial Test Performance.....                | 46 |
| 2.2.2.2.Final Test Performance.....                  | 47 |
| 2.2.2.3.Forgetting Rate and the Testing Effect ..... | 50 |
| 2.2.3. Discussion.....                               | 51 |
| 2.3.Experiment 3.....                                | 53 |
| 2.3.1. Method.....                                   | 53 |
| 2.3.1.1.Participants.....                            | 53 |
| 2.3.1.2.Materials.....                               | 54 |
| 2.3.1.3.Design.....                                  | 55 |
| 2.3.1.4.Procedure.....                               | 55 |
| 2.3.2. Results.....                                  | 56 |
| 2.3.3. Discussion.....                               | 59 |
| 2.4.Experiment 4.....                                | 60 |
| 2.4.1. Method.....                                   | 62 |
| 2.4.1.1.Participants.....                            | 62 |
| 2.4.1.2.Materials.....                               | 62 |
| 2.4.1.3.Design.....                                  | 63 |
| 2.4.1.4.Procedure.....                               | 64 |
| 2.4.2. Results.....                                  | 65 |
| 2.4.2.1.Scoring.....                                 | 65 |
| 2.4.2.2.Initial Test Performance .....               | 65 |
| 2.4.2.3.Final Test Performance.....                  | 65 |
| 2.4.2.4.Initial and Final Performance.....           | 67 |

|  |            |
|--|------------|
| 2.4.3. Discussion.....                   | 68         |
| <b>General Discussion.....</b>           | <b>70</b>  |
| 2.5. Transfer of the Testing Effect..... | 70         |
| 2.6. Theoretical Implications.....       | 75         |
| 2.7. Practical Implications.....         | 80         |
| 2.8. Future Directions.....              | 84         |
| <b>References.....</b>                   | <b>87</b>  |
| <b>Appendix A.....</b>                   | <b>95</b>  |
| <b>Appendix B.....</b>                   | <b>101</b> |
| <b>Appendix C.....</b>                   | <b>109</b> |
| <b>Appendix D.....</b>                   | <b>110</b> |
| <b>Appendix E.....</b>                   | <b>120</b> |
| <b>Appendix F.....</b>                   | <b>129</b> |
| <b>Appendix G.....</b>                   | <b>144</b> |
| <b>Appendix H.....</b>                   | <b>163</b> |

# List of Tables

|  |           |
|--|-----------|
| <b>Table 1 - Percentage of Questions Answered Correctly on the Final Test as A Function of</b> |           |
| <b>Initial Task Type (N = 41) .....</b>  | <b>66</b> |



## **List of Figures**

|  |           |
|--|-----------|
| <b>Figure 1 - Different Domains and Dimensions of Transfer Changing in the Continuum of Near and Far Transfer as Proposed By Barnett and Ceci (2002) .....</b>   | <b>18</b> |
| <b>Figure 2 - Mean Percentage of Correct Answers as A Function of the Type of Questions Asked on the Final Test in the Laboratory Study of Experiment 1.....</b> | <b>40</b> |
| <b>Figure 3 - The Design of Experiment 2.....</b>  | <b>45</b> |
| <b>Figure 4 - The Interaction between the Type of Questions (i.e., factual vs. conceptual) and 1 Week Interval.....</b>  | <b>48</b> |
| <b>Figure 5 - The Transfer of the Testing Effect as A Function of Question Type (I.E., Factual Vs. Conceptual) and the Encoding Type.....</b>                    | <b>49</b> |
| <b>Figure 6 - Forgetting Rate as A Function of the Question Type (i.e., Factual vs. Conceptual) Throughout 1 Week Delay.....</b>                                 | <b>51</b> |
| <b>Figure 7 - The Construction of Questions of the Final “Practice” Test.....</b>  | <b>55</b> |
| <b>Figure 8 - The Transfer of the Testing Effect as A Function of Question Type and the Type of Encoding Task. ....</b>  | <b>57</b> |
| <b>Figure 9 - The Design of Experiment 4.....</b>  | <b>64</b> |
| <b>Figure 10 - The Performance on the Final Test as A Function of the Match between Initial Task and the Final Task.....</b>                                     | <b>67</b> |
| <b>Figure 11 - The Forgetting Rate Throughout 1 Week as A Function of the Question Format. ....</b>  | <b>68</b> |

# 1. Introduction

In educational settings, the common objective shared by both educators and students is the acquisition of knowledge, which is evaluated by tests in which students have to give evidence that they remember the taught materials in the course (Bloom, 1956). This traditional aspect of tests (i.e., evaluating how well students have learned the course material) has made educators and students disregard another function of tests for many years: tests can also be used as a learning device. Although tests are important for evaluating and measuring learning, tests also improve students' memory for content (Modigliani & Hedges, 1987). Taking a test on the material produces better retention than spending time restudying the material. The memory gain produced by tests is called the *testing effect* (see Roediger & Karpicke, 2006a, for a review). The present study aimed to evaluate the effectiveness of the testing effect by using different types of transfer tasks and to highlight the underlying mechanisms of the testing effect according to relevant theories.

## 1.1. The Testing Effect: The Paradigm

A great deal of research has shown that testing produces benefits to learning (e.g., Glover, 1989; McDaniel & Masson, 1985; Roediger & Karpicke, 2006a). Although designs (e.g., within-subject vs. between-subject; laboratory vs. classroom etc.) and materials (e.g., word lists vs. prose passages; recognition vs. recall tests) used in these studies show variation, the general procedure is very similar. A fairly traditional paradigm for studying testing effects is first reviewed here.

Experiments examining testing effects start with the presentation of to-be-learned material which can be word lists and prose passages. In this phase (i.e., encoding phase), participants are asked to study the material for later memory tests. This type of learning is called intentional learning in which subjects are told that their memory of the experimental material will be tested later (Bereiter & Scardamalia, 1989). Following the first phase, participants are given either an intervening study trial, in which they restudy the materials (i.e., a study-only condition), or a test trial, in which participants are given a test on the presented materials (i.e., testing condition). There are different variations of testing trials. Participants can be asked to recite as much of the material as they can recollect or they can be given a multiple-choice test where they are required to choose the correct answer among given options. The former is known as a recall task and the latter is known as a recognition task (Watkins & Gardiner, 1979). These intervening trials are followed by a final memory test to assess the amount of learning produced by intervening trials. The final memory test usually is called a final criterion test. The final memory test can be given right after the intervening trials or after a delay, such as 2 days, 1 week or 1 month later. Memory accuracy on the study-only condition and the testing condition on the final test is compared. The well-established finding is that participants perform better on the final memory test following intervening testing trials. In other words, taking a test on to-be-learned material during encoding promotes learning. In the following section, a review of some of the studies which investigated the testing effect in a variety of situations is provided.

## **1.2. The Testing Effect in the Laboratory and in the Classroom**

Although the initial steps investigating the facilitative effect of testing on memory performance go back to the beginning of 20<sup>th</sup> century (e.g., Gates, 1917), apart from a handful of studies in the 1980s and 1990s (e.g., Glover, 1989; McDaniel & Masson, 1985), the topic

appears to have been nearly forgotten by researchers until the beginning of 21<sup>st</sup> century. To emphasize the lack of interest in the topic, Glover (1989) published an article on the testing effect with the title “The ‘Testing’ Phenomenon: Not Gone but Nearly Forgotten.” In the following years, newly found interest in the testing effect has flourished among cognitive psychologists. In this section, a chronological review of the most important studies conducted on the testing effect will be presented.

Although there was other research which demonstrated the benefits of taking a test to promote retention before Gates (1917) (e.g., Abbot, 1909), Gates was the first researcher to conduct a systematic study on the testing effect. His main argument was “poor memories may prove to be due to poor methods of study” (p. 2). Gates (1917) argued that self-testing should provide a good method of study. In his study, he tested students from different grades ranging from the first grade to the eighth grade. He used two types of materials: nonsense syllables and brief biographies. He also manipulated the testing time. After students read the material (i.e., encoding phase), they were asked to look away and recall the material to themselves (i.e., self-testing). Afterwards, students were given a final memory test in which they were asked to write down as many items as possible. Students were retested after 4 hours. The results showed that reciting information during learning/encoding improved retention on subsequent tests. He concluded that self-testing is a valuable way for promoting learning. Gates’ findings were replicated by numerous studies both in the laboratory and classroom settings, utilizing a variety of materials, such as lists of words and prose, and different types of tests, such as recall and recognition tasks even without the corrective feedback (Roediger & Karpicke, 2006a).

After establishment of the testing effect, much effort has been put into investigating the ways of improving the testing effect. The first finding which is worth mentioning is the effect of

repeated retrieval attempts on memory which can be summarized as follows: taking a test is good for retention of knowledge, and taking multiple tests is even better. For example, Roediger and Karpicke (2006b), in Experiment 2, assessed the testing effect with educationally relevant materials, which were passages selected from the reading comprehension sections of the preparation book for the TOEFL. After reading the passages, participants either reread the passages (read-only condition) or took one or three immediate recall tests depending on the condition they were assigned to. There were three conditions (S = Study, T = Test): repeated study condition (SSSS), single test condition (SSST) and repeated test condition (STTT). Participants, then, took a final memory test (free recall) 5 minutes or 1 week later. Results revealed that as the number of intervening tests increased, participants recalled more information on the retention test and their forgetting rates decreased. The main conclusion inferred from the study is that testing is a valuable tool for learning, but repeated testing is even more valuable.

Although earlier studies reported that the testing effect was limited to items tested (e.g., Runquist, 1983), later research showed that the testing effect is observed not only for tested information but also nontested but related information. Chan, McDermott and Roediger (2006) examined how testing influences retention of nontested information. In the study, participants studied an article, for example, on the toucan bird. The following sentence is taken from the article: "Toucans cannot make the holes themselves due to their soft bills; instead they sleep in tree holes which are made by woodpeckers." After reading the article, they were either given a test including 40 short-answer questions or asked to reread the article. For example, one of the questions in the initial/intervening test was as follows: "Where do toucans sleep at night?" On the second day, participants took a final test with 40 questions; some of them were the same questions that were tested before (old-questions) and some of them were the questions that were

not tested but were conceptually related to the initially tested information (nontested-related). For example, the participants were asked the following question once again: “Where do toucans sleep at night?” Or they were asked the following question as nontested relevant question: “What other bird species is the toucan related to?” The answer of the first question was “tree holes” and the answer of the second question was “woodpeckers”. These two answers were closely related to each other as indicated in the article: toucans cannot make the holes themselves due to their soft bills, instead they sleep in tree holes which are made by woodpeckers. The results showed that initial testing is beneficial not only for tested information but also for nontested-relevant information. Participants who were given additional study time (a read-only condition) did not demonstrated such a facilitative effect of testing.

Tests not only improve memory but also protect memory from retroactive interference which occurs when recently learned information blocks the retrieval of previously learned information (Baddeley & Dale, 1966). Halamish and Bjork (2010) investigated whether tests can overcome retroactive interference in a cued-recall paradigm. In the study, there were four phases: (1) the study phase where the participants studied the word pairs (e.g., *cut: knife*), (2) the practice phase where the participants were provided either with the first word and the fragments of the second word of the pair (e.g., *cut: \_\_i\_\_e*) (i.e., testing condition) or with the word pair in an intact form (e.g., study condition), (3) the distractor phase where the participants were presented with the same word pair as in the study phase (*cut: knife*) or with their distractor counterparts, such as *cut: slice*. Please note that *cut: \_\_i\_\_e* can be interpreted as *cut: knife* or *cut: slice*. In the fourth phase, final test phase, the participants were given a cued-recall test where the first word of the pair was presented intact and the fragment of the second word was presented (*cut: \_\_i\_\_e*) and the participants were asked to complete the fragments. As predicted, Halamish and Bjork

(2010) found retroactive interference: recall rates decreased from practice to final test in the interference condition. Repeated testing did not eliminate the retroactive interference but reduced the interference more than did restudying. This study reveals that tests can protect against the deteriorative effect of retroactive interference.

The testing effect has been studied in actual classroom settings, although these studies are few in number (Butler & Roediger, 2007; McDaniel, Anderson, Derbish & Morissette, 2007). For example, McDaniel et al. (2007) experimentally examined the testing effect for the curriculum covered throughout the semester in a college course. Students took weekly quizzes about the target material or were given some additional reading about the target material as an exposure control. At the end of the semester, they were given a cumulative multiple choice test. The results showed that quizzing improved performance on the final test. Such an effect was not observed in the additional reading condition.

Tests facilitate learning not only in the form of intervening tasks but also in the form of pre-study tasks. Little and Bjork (2010) conducted a study where the participants were given the test (i.e., multiple-choice) before or after studying the experimental materials. Results showed that taking a pre-test improved the performance on a later cued-recall task. Moreover, taking a pre-test did not increase the likelihood of the intrusions even when the participants chose the wrong answer on a pre-test. This study draws attention to the role of test as tools for potentiating learning.

As can be seen in the review above, the testing effect is a robust phenomenon. Accumulated evidence on the testing effect directed researchers to explore the underlying mechanisms of the effect which, in turn, paved the way for the proposal of different theoretical

orientations. What are the possible mechanisms underlying the testing effect? In the following section, theoretical explanations of the testing effect are reviewed.

### **1.3. Theories of the Testing Effect**

Different explanations have been put forth to explain why the testing effect exists. For a long time, the overlearning/exposure hypothesis dominated the testing effect literature. However, recently, most researchers adhere to two other frameworks explaining the testing effect: the retrieval hypothesis and the transfer-appropriate processing view. Although the present study aimed to compare and contrast the retrieval hypothesis and the transfer-appropriate processing view, the overlearning/exposure hypothesis is also reviewed in this section to give the readers an idea about the development of theoretical orientations of testing effects.

#### **1.3.1. Overlearning/Exposure Hypothesis**

The overlearning/exposure hypothesis is the first theoretical explanation proposed for the testing effect. Thompson, Wenger and Bartling (1978) proposed that the testing effect is the result of re-presentation (in the form of tests) of the material studied during the initial encoding phase. In other words, the testing effect is a result of rote memorization of to-be-tested-information through additional exposure. The early adherents of this hypothesis argued that the obtained testing effects in the literature arise from the fact that experiments (e.g., Duchastel, 1979) investigating testing effects compare conditions in which students study materials, complete a filler task, then take the final test to a condition in which subjects study, retrieve information, then take the final test. Hence this type of design gives advantages to material learned in the initial testing condition in terms of extra exposure of the material since the subjects in the testing condition spend more total time on the to-be-learned material. In other words, successful recollection of the material during the initial learning period provides additional



exposure to the tested material, and the extra exposure is responsible for the testing effect. However, throughout the years, with proper experimental control, researchers still have obtained the testing effect: after equating exposure time to the to-be-learned material in study-only and testing conditions, the testing effect was still present (e.g., Nungester & Duchastel, 1982). Later on, the adherents of this theory held that recalling or reciting information facilitates later recollection of that information. The reason that reading or studying information yields poorer retention than reciting information is because reciting information improves retention through different mechanisms than does the presentation of information (Yuille, 1973). Although participants are exposed to all of the to-be-remembered materials on study/read trials, at the time of initial testing they recall only a subset of the items (Carrier & Pashler, 1992). Hence the list that the participants recall at the initial test is a shorter version of the original material and it is composed of best-learned items, which gives them an advantage on the final test. On the other hand, although the participants can get exposed to the same well-learned items, in the study-only condition, they also have an opportunity to study less well-learned items, which might be the source of greater interference effect on the final test, which in turn, leads to poorer memory (Toppino & Cohen, 2009). In a series of experiments Thompson et al. (1978) proposed that overlearning due to extra exposure of best-learned items was responsible for the testing effect. In Experiment 1, they set up three experimental conditions. One condition involved multiple presentations of experimental material in which subjects were exposed to four presentation trials of the same word list. Another condition was a multiple recall condition in which subjects were presented with a word list in the first trial and were required to recall the words in the following three trials. The last condition was a recall-plus-re-presentation condition which was the same as the multiple recall condition, except that un-recalled words were re-presented on each recall trial,

which granted subjects extra exposure of the unrecalled items (i.e., an overlearning manipulation; this is essentially recall with feedback). Recall-plus-representation yielded the highest memory performance. Additional analysis of data showed that an average subject recalled a subset of items and then practiced this subset in the following trials. In other words, the subject overlearned this subset of items which helped the subject locate the items in memory. Based on the results, Thompson et al. (1978) stated that the testing effect is not the result of retrieval processes but a result of overlearning a subset of items. However, this study essentially examined the testing effect with feedback, and it fails to explain how the best-learned items are being overlearned (for example, overlearning typically involves multiple trials, and it is possible to see the testing effect with just one testing trial). Moreover, the study fails to provide any explanation why testing yields overlearning but multiple exposures through study condition do not lead to overlearning. In this regard, this study does not seem to shed light on the mechanisms underlying the testing effect.

The main challenge to this hypothesis comes from studies which showed long-term testing effects, in which testing produced a better long-term retention rate compared to a study-only condition (e.g., Wheeler, Ewers, & Buonanno, 2003). In the immediate final memory test, studied-only items were recalled/recognized better than tested items. Apparently, overlearning is pronounced for only delayed tests, and the overlearning hypothesis cannot provide an explanation for this interaction. Another problem for this account comes from the study conducted Chan et al. (2006) as reviewed above. In this study, a testing effect was observed not only for the tested items but also non-tested related items. In a series of experiments, the authors investigated how initial testing affects later memory for prose material which was not tested but which was conceptually related to the tested material. The study showed that initial testing not

only enhanced the retention of the tested material but also enhanced recall of untested but relevant material. This effect was not found in conditions where participants were given additional study time and no testing opportunities. The overlearning hypothesis fails to provide any explanation, because this study eliminates the factor of overexposure (i.e., untested relevant information was not presented in the initial test.). To sum up, the overlearning hypothesis, alone, is not enough to explain the testing effect. The failure to provide convincing explanation with proper scientific evidence decreased the credibility of this framework. It seems that the testing effect is not a product of overlearning, mere additional exposure, and it reflects more than rote learning. The accumulating evidence guided researchers to develop more comprehensive frameworks explaining the testing effect: the retrieval hypothesis and the transfer-appropriate processing framework. Therefore, it is these two theories that the current study will emphasize in analysis and discussion.

### **1.3.2. Retrieval Hypothesis**

The adherents of this hypothesis argue that the testing effect is not a function of overexposure of the to-be-learned material during encoding but is a function of retrieval mechanisms during encoding (Carpenter, 2009). There are two aspects of this hypothesis which are not necessarily independent from each other as is discussed below. The first one is the retrieval effort: at the time of encoding, the relative difficulty of retrieving information rather than re-studying it strengthens memory traces (Karpicke & Roediger, 2006b). The challenges experienced by learners during encoding are responsible for superior retention. This situation is also described as “desirable difficulties” by Bjork (1994). Desirable difficulties refer to learning conditions with difficulty which leads to more flexible and enduring learning. As the difficulty of the encoding process increases, the mental elaboration on the material also increases. The well-

known example of desirable difficulties is *generation effect* which refers to the fact that items generated by the participants are remembered better than read items (Hirshman & Bjork, 1988). For example, Begg, Snider, Foley and Goddard (1989) reported that the subjects remember items they complete from fragments (el\_ph\_n\_) better than the items they read (elephant). Item generation is considered to be more difficult than reading. Superior retrieval due to generation is attributed to the relative difficulty of the act of generation as compared to simply reading the words. Some readers might think that the generation effect is a version of the testing effect. However, although they look similar, the processes underlying the generation effect and the testing effect are different: the generation effect relies on incidental learning (i.e., it does not necessitate intentional retrieval of previously learned items) whereas the testing effect relies on intentional retrieval where participants are required to retrieve material that they had learn the previously in the experiment (Karpicke & Zaromb, 2010). The only thing they have in common is the notion that retrieval difficulty is responsible for later retrieval. In this regard, using tests as encoding creates learning conditions with optimal difficulty for the learners as argued by Roediger and Karpicke (2006a). This argument has been supported by numerous studies. For example, Gardiner, Craik and Bleasdale (1973) investigated the relationship between the amount of time to answer general knowledge questions and the retention of the answers. The students took a test on general knowledge. At the time of the initial test, the researchers measured the amount of time it took them to answer the questions. After the session, the students were given a final recall test on the answers. The results showed that the longer it took to answer a question, the more likely it was recalled correctly on the final test. The researcher argued that the amount of time spent on questions was a function of question difficulty, hence an indication of greater retrieval effort, which resulted in improved retention. Similar findings were reported by other

studies (e.g., Auble & Franks, 1978). However, it should be noted that personal characteristics of participants (e.g. working memory capacity, personal interests) might have affected the time spent on the question, yet personal characteristics were not in the scope of the article. Moreover, one can argue that the amount of time spent on questions might be responsible for the superior retrieval not because of a function of difficulty but because of exposure time since participants spent more time on these items. Therefore, the operationalization of difficulty in terms of time does not seem appropriate for experimental purposes. Fortunately, recently Pyc and Rawson (2009) published a study which manipulated the difficulty in a different way: interstimulus interval (the number of items between each practice trial). Based on previous research, Pyc and Rawson (2009) predicted that a longer interstimulus interval (ISI) would lead to more difficult retrieval yet to better final memory performance. These predictions are consistent with what would be predicted by the retrieval hypothesis. What they found supported the predictions of this framework: longer ISIs led to more difficult retrieval (as measured by response time) yet led to higher levels of final test performance. Additionally, increasing the number of times a subject had to correctly retrieve an item facilitated initial retrieval yet diminished the testing effect. These studies support the prediction that difficulty of retrieval at the time of encoding is one of the mechanisms responsible for the testing effect. Moreover, it can be argued that retrieval activity in the initial test as compared to rereading the material requires individuals to engage in more elaborative encoding processes which results in superior retention.

As indicated above, there are two aspects of the retrieval hypothesis. The first one is related to retrieval difficulty. The second aspect concerns the changing of memory traces during retrieval attempts (Carpenter, 2009). It states that a retrieval attempt at the time of encoding changes memory traces in a way that it produces an elaboration of existing memory traces and

proliferates retrieval routes, which increases the variability of encoded information (i.e., encoding variability). The increased variability of encoded information provides individuals with a vast array of cues which can be used to retrieve information at a later time (McDaniel & Masson, 1985). This aspect is also in line with the spreading activation theories of memory (Collins & Loftus, 1975) which states that retrieval activity affects not only the item of interest but also other related concepts which, in turn, activates the semantic network with multiple pathways leading to the item of interest. For example, a subject is given a word (e.g., *butterfly*) to learn for a later memory test. Instead of simply rereading the word pair, trying to recall the word activates not only the word *butterfly*, but also all related concepts, such as *flower*, *fly* and *caterpillar*. Therefore, the future recall of *butterfly* is facilitated via the activation of these related concepts since even when the individual fails to recall *butterfly*, s/he can rely on other memory cues (*flower*, *fly* and *caterpillar*) to reach the target word (*butterfly*). Reorganization of memory traces at the time of testing was also acknowledged by Bloom (1956). He argued that any testing situation requires reorganization of the problem to supply appropriate signals and cues related to the problem. Hence, after taking an initial test, an individual is more likely to find the appropriate signal and cues to solve a problem on a final test. Encoding variability was supported by numerous empirical researches (e.g., McFarland, Rhodes, & Frey, 1979). For example, spacing of the initial test promotes better retention because every retrieval attempt augments the features of the memory trace itself which provides more cues to retrieve the correct answer on a later test (Karpicke & Roediger, 2006a; Logan & Balota, 2008).

This hypothesis is accepted widely among scholars. However, before accepting this explanation, more studies are needed with proper manipulations of retrieval difficulty and encoding variability which is one of the aims of this study.

### **1.3.3. Transfer-Appropriate Processing View**

Transfer appropriate processing is one of the longer-standing theories in the field of cognitive psychology (Morris, Bransford & Franks, 1977). It posits that the value of a particular encoding or acquisition strategy can only be evaluated with regards to the context of the particular retention test used to evaluate learning. Learning is optimized when the encoding processes and retrieval processes are similar (Lee, 1988). This framework relies on the principle of encoding specificity which states that memory is enhanced when information presented at encoding is also presented at retrieval (Tulving & Thompson, 1973). Godden and Baddeley (1975) provided a widely-cited study exploring the encoding specificity principle. In the study, participants learned lists of words either on land or under the water. In the retention test, they were asked to recall the words either in the same setting or in the opposite setting. The results showed that the participants who were tested in the same condition recalled more words than in the opposite condition.

Several other studies demonstrated the mnemonic advantage of mental activities during encoding that will be needed at retrieval. For example, Morris et al. (1977) reported that semantic acquisition yielded superior performance in a recognition test than did rhyme acquisition whereas rhyme acquisition was superior in a rhymic recognition test than did semantic acquisition. Transfer-appropriate processing approach regards the testing effect as a result of similarity between initial and final test: an initial memory test taken during encoding, as compared to reading the material, improves retention due to the fact that retention is also measured by another memory test (Thomas & McDaniel, 2007). In simpler terms, the initial test shows more resemblance to the final test than does restudying alone (Carpenter, 2009). Encoding processes employed by reading materials are fundamentally different than the retrieval

processes employed by taking a test (i.e., there is a mismatch between encoding process and retrieval process). On the other hand, encoding processes used while taking a test are conceptually similar to retrieval processes required by a test (i.e., there is a match between the encoding process and the retrieval process). As noted by Roediger and Karpicke (2006a), the transfer-appropriate processing view offers a very intuitive explanation for the testing effect. However, this approach has a very specific prediction which has yet to receive empirical support. Transfer-appropriate processing predicts that performance on a final test should be best when the test has the same format as initial test. Several studies have investigated the effect of format on the testing effect (e.g., Wenger, Thompson & Bartling, 1980) yet very few studies systematically addressed this issue (e.g., Kang, McDermott & Roediger, 2007). Earlier studies conducted to investigate this issue reported different findings on this issue. For example, Weger et al. (1980) reported that an initial recall test improves recognition memory on later memory tests. Runquits (1983, Exp. 1) reported that intervening recognition tests improved later recall performance. Nungester and Duchatsel (1982) showed that the testing effect induced by an initial recognition or recall test was still observed when the formats reversed in the final tests. These studies, however, lack the experimental control, such as factorial manipulation of the formats of both initial and the final test. Indeed, very few studies have manipulated both formats. For example, Kang et al. (2007) manipulated the formats of both the intervening and final tests (i.e., multiple-choice and short-answer questions). The results of their Experiment 1 produced an interesting picture: taking multiple-choice tests produced superior testing effects regardless of the format of the final test, and initial short-answer tests produced better performance than a study-only (i.e., control condition) only when the final test was a short-answer format. However, there is another study which showed that the testing effect was strongest when the initial test was a free recall



task regardless of the final test (Carpenter & DeLosh, 2006). These findings will be discussed in the following section in detail. For now, for the purposes of the current section, it is enough to postulate that the predictions of transfer appropriate processing view seem to fail to receive much supporting evidence from empirical studies. However, as proposed by Roediger and Karpicke (2006a), this statement needs confirmation through further studies due to the relatively few numbers of studies and lack of consensus on the issue.

#### **1.4. Transfer of Learning**

Although the number of studies investigating the testing effect and factors affecting the testing effect has accumulated gradually, one topic, interestingly, has not received much attention from researchers: transfer of learning gained via tests. The transfer of learning obtained with testing should give some ideas about the depth of learning occurring in the testing effect. Just how powerful is it? This section reviews the transfer of learning literature. Note that since the literature on the transfer of learning is very extensive, the following review focuses on studies which are most relevant to the purposes of the present paper.

Transfer of learning has been a very popular research topic in psychology (see Barnett & Ceci, 2002). In simplest terms, it can be described as "...through experience in learning we get better at learning" (Larkin, 1989, p. 283). Transfer of learning occurs in different forms. One can apply old information into new situations to solve a novel problem or one can use old knowledge in a new situation to learn new knowledge (Larkin, 1989). Transfer of learning also takes place when previously learned knowledge or skills facilitates newly learned knowledge and skills (Simons, 1999). Transfer of learning has important implications for basic and applied research. Outcomes of transfer research are important in terms of both money and time invested in education. Students are expected to translate knowledge that they learn in school to daily life.

Therefore, formal schooling should indoctrinate general skills which should transfer beyond academia and help students become a productive members of society (Barnett & Ceci, 2002). For example, it is important to know if an engineering student will be able to use his knowledge from a course on fluid dynamics in the field while troubleshooting a water pipeline of a city. Or will a clinical psychology student be able to apply therapy techniques that she learned in school on her patients?

Although we can talk about different types of transfer, such as transfer of skills or training, the scope of the present study is on transfer of memory gains obtained with testing. Transfer of the testing effect deserves attention due to the fact that although the testing effect has been reported to be a very powerful tool for learning, it is an open question whether testing promotes deeper levels of learning which can be evaluated through transfer of learning. The research question can be rephrased in the following way: how efficiently can learners extract information they gain through tests and use it in a different context? In this regard, we have yet to assess (1) whether memory boosts obtained through tests can be long-lasting and transferable into new contexts and (2) how we can utilize testing effects to promote effective transfer. The present study attempted to answer these questions.

#### **1.4.1. Dimensions of the Transfer of Learning**

Why is transfer of learning important? Can we transfer what we learn? How similar does the learning context have to be to boost transfer? These are among the questions that researchers have attempted to answer, and their general conclusions are reviewed here.

What is transfer? This question has been answered in different ways. It might refer to the transfer of performance or knowledge. For example, Woodworth and Schlosberg (1954) defined it as “the carrying over of an act or way of acting from one performance to another” (p. 734). On

the other hand Bransford, Brown and Cocking (1999) defined it as “the ability to extend what has been learned in one context to new contexts” (p. 39). Barnett and Ceci (2002) suggested that there are different domains and dimensions of transfer changing in the continuum of near and far transfer (Figure 1). They identified six different domains of transfer: knowledge domain, physical context, temporal context, functional context, social context, and modality. Knowledge domain refers to transfer of learning across different subjects. Physical context refers to correspondence between the learning and testing environment. Temporal context refers to time elapsed between learning and testing. Functional context is the match between applicability of learned information at the time of learning and testing. Social context refers to whether the task is learned and tested alone or in collaboration with other people. The last dimension is called modality which refers to correspondence between format of to-be-learned information during learning and testing. Since the present study focuses on knowledge domain, temporal context, and modality, these three domains are discussed in detail with relevance to the testing effect.

**Figure 1. Different domains and dimensions of transfer changing in the continuum of near and far transfer as proposed by Barnett and Ceci (2002).**

| <b>B Context: When and where transferred from and to</b> |                           |   |                                   |                                     |                          |
|--|---------------------------|---|-----------------------------------|-------------------------------------|--------------------------|
|  | Near ← → Far              |   |                                   |                                     |                          |
| <b>Knowledge domain</b>                                  | Mouse vs. rat             | Biology vs. botany                      | Biology vs. economics             | Science vs. history                 | Science vs. art          |
| <b>Physical context</b>                                  | Same room at school       | Different room at school                | School vs. research lab           | School vs. home                     | School vs. the beach     |
| <b>Temporal context</b>                                  | Same session              | Next day                                | Weeks later                       | Months later                        | Years later              |
| <b>Functional context</b>                                | Both clearly academic     | Both academic but one nonevaluative     | Academic vs. filling in tax forms | Academic vs. informal questionnaire | Academic vs. at play     |
| <b>Social context</b>                                    | Both individual           | Individual vs. pair                     | Individual vs. small group        | Individual vs. large group          | Individual vs. society   |
| <b>Modality</b>  | Both written, same format | Both written, multiple choice vs. essay | Book learning vs. oral exam       | Lecture vs. wine tasting            | Lecture vs. wood carving |

*Knowledge domain* is the first dimension of transfer context where the individual transfers what he learns from one content area to another. For example, an episode cited by Bloom (1956) demonstrates this type of transfer. A professor asks a class “What would you find if you dug a hole in the Earth?” He gets no response. Then he repeats the question “What is the state of the center of the earth?” Then the class replies in unison, “Igneous fusion.” Bloom (1956) argued this episode illustrates the rote nature of some knowledge learning which seemingly prevents transfer of learning because students could not access the memory trace already available in memory. Bloom (1956) argues that rote learning is different from “true knowledge” (p. 29). In Bloom’s terminology, true knowledge seems to refer higher level knowledge which can be utilized in new situations and in a form which is very different from where it was originally learned. The transfer in the knowledge domain can be stretched from near transfer to far transfer. Far transfer is a function of the similarity between learning context and testing context. As the similarity diminishes, the transfer is considered to be far transfer.

The other dimension of transfer context is *temporal context*, which is related to the endurance or retention of learning. It reflects the time passed between learning and testing phases (e.g., minutes, hours, weeks, years). To justify effort invested in education, one should assume that one of the objectives of education is to equip students with knowledge which would last for several years after training. As in the knowledge domain, transfer in the temporal domain can be stretched from near transfer to far transfer. In this case, far transfer is a function of the time passed between the learning context and testing context. As the time elapsed between learning and testing phase increases, the transfer is considered to be far transfer (Barnett & Ceci, 2002).

The final dimension of transfer context is *modality*, which concerns the procedures or format of learning and testing. As Barnett and Ceci (2002) suggested, this transfer context can be

at the micro or macro level which corresponds to near and far transfer, respectively. As the correspondence of tasks between learning and testing phase decrease, the transfer is considered to be far. When both learning and testing are in the same format (e.g., verbal or written), the transfer is near whereas when learning and testing are in different formats, then the transfer is considered to be far. For example, when an engineering student learns fluid dynamics in the class in a written and verbal format but uses this information to constructing water pipe lines, this is considered to be an example of far transfer. On the other hand, learning via written format and testing via written format is an example of near transfer (Barnett & Ceci, 2002).

This thesis offers a novel approach for assessing the testing effect: the transfer of the testing effect in the three domains summarized above. In the following section, the implications of this approach are discussed at length.

## **1.5. Transfer of the Testing Effect.**

Although it has been repeatedly demonstrated that the testing effect is robust across several materials and across several settings (i.e., laboratory and classroom), not many studies have investigated transfer of the testing effect, especially as it relates to transfer in the knowledge domain. Studies have generally focused on modality (i.e., transfer across different test formats) and/or on temporality (i.e., retention of learning over time).

### **1.5.1. Transfer in the Temporal Domain**

The transfer of the testing effect in the temporal domain is a well-established fact by both laboratory and field research. Numerous studies reported that the testing effect is observed in delayed final tests. For example, Carrier and Pashler (1992) examined testing effects on a paired-associate task. Participants were presented with word pairs one at a time. After the initial presentation, participants were either presented with the pair again (i.e., the study condition) or

with the first half of the pair (i.e., the test condition). In the testing condition, participants were required to recall the other half of the pair. The researchers equated the total time in the study and test conditions. The results revealed a superior performance in the test condition on a delayed memory test. Furthermore, the transfer of testing effects in the temporal domain is not limited to word lists. For example, Roediger and Karpicke (2006b) examined the transfer of testing effects with educationally relevant materials. In two experiments, participants studied prose passages covering a single topic. They manipulated learning conditions (restudy vs. test) and delay of the final test (5 minutes, 2 days or 1 week). After reading the passages, the participants in the restudy condition reread the passages whereas the participants in the test condition were asked to write down as much of the material as they could remember from the passages (i.e., free recall). The final recall test was given after 5 minutes or after 2 days or after 1 week. The results revealed that although the restudy condition resulted in a higher proportion of recalled units on the immediate final test (i.e., 5 minutes delayed), after a delay (2 days and 1 week), the memory performance of participants in the test condition was superior to their counterparts in restudy condition.

As indicated above, the testing effect seems to promote long-term retention not only in the laboratory but also in the classroom. For example, McDaniel et al. (2007) experimentally examined the testing effect for the content presented throughout the semester in a college course. Students took weekly quizzes about the target material or were given some additional reading about the target material as an exposure control. At the end of the semester, they were given a cumulative multiple choice test. The results showed that quizzing improved performance on the final test. Such an effect was not observed in the additional reading condition. Similar results were obtained in other studies as well (e.g., Nungester & Duchastel, 1982). These studies show

that transfer of testing effects in the temporal domain observed in strict laboratory settings is also observable in settings outside the laboratory.

One of the objectives of the present study is to replicate and extend these findings in the temporal domain, exploring whether the testing effect can be transferred across memory tests with relatively long delays (i.e., 1 week and 12 weeks). Based on the literature reviewed above, it is expected to find long-term facilitative effects of the testing effect.

### **1.5.2. Transfer in the Modality Domain**

The second domain of interest is the transfer in the modality domain, namely the transfer of testing effects across different test formats. In other words, the question of interest is whether the testing effect observed is independent of the nature of the initial test and the final test. Contrary to research in the temporal domain, research on modality has yielded contradictory results. Some studies reported that the testing effect occurs as long as the final criteria test involves recall (e.g., Chan & McDermott, 2007) pointing out that transfer occurs only when the final test is a recall task. On the other hand, some researchers have demonstrated that taking an initial free recall test produced the best testing effect regardless of the format of the final test (e.g., Butler & Roediger, 2007; Carpenter, Pashler & Vul, 2006; Glover, 1989) indicating that transfer is independent of final test but is a function of the initial test. Because of the contradictory results, it is crucial to review these studies and discuss possible reasons for such discrepancy.

One of the earliest studies on transfer of testing effects across different test formats was conducted by Duchastel (1981). Before studying the to-be-remembered materials, the students were informed that a retention test would be administered to assess their memory of the material. After the study phase, the students took an either initial short-answer or multiple-choice test on a

prose passage, and then took a final short-answer test 2 weeks later. The planned comparison between short-answer and multiple-choice tests showed that short-answer tests produced the best final retention (i.e., a greater testing effect). Similarly, in a very similar study, Butler and Roediger (2007) reported that short answer questions generated the greatest testing effect. In fact, the testing effect was found only with short answer questions; multiple choice questions did not produce any testing effect. The difference between the studies might be caused by the materials used in the studies or the retention interval; the former one utilized 2 weeks and the later one utilized 1 month. However, the findings of these studies should be interpreted with caution: they did not factorially manipulate the formats of the tests in the initial and the final test sessions. It might be argued that the reason for the advantage of the short answer questions in the final test might be the similarity of the formats in the initial and the final test. Hence, the superiority of the testing effect produced by an initial recall test might be an artifact of similar processes employed in the initial test and the final test. To be able to infer that the initial free-recall test yields superior transfer, it is essential to manipulate the formats of both the initial test and the final tests. Glover (1989), in a more elaborate study which factorially manipulated the formats of the questions in the initial and the final test, overcame the problems introduced by Duchastel (1981). In the study (Experiment 4), two days after reading a 300-word essay on a single topic, participants in four different conditions were given different formats of an initial test: free-recall, cued-recall, or recognition test. Four days later, they took different types of final criterion tests. The results revealed that free-recall tests improved performance more than any other test. However, the findings should be taken cautiously once again. The most important drawback of this study is the differential amounts of testing: on the cued recall tests, 12 idea units were tested whereas on the recognition test, 6 idea units were tested. Also, Glover (1989)



did not control for the exposure time: participants in the study condition were not given extra time to equate the time given to their counterparts in the testing condition. A study conducted by Carpenter and DeLosh (2006) overcame the problem of the studies reviewed above. They manipulated the formats of initial and final tests (i.e., recognition, cued recall and free recall) in a way that in the final test the format of some of the questions matched the format of the initial test (e.g., recognition-recognition) or not (e.g., recognition-free recall). They also controlled the exposure time by giving the participants in the study-only condition extra time to restudy. The final test performance was the best for the items which were given a free recall initial test. The format of the final test did not affect the magnitude of the testing effect.

On the other hand, as reviewed in the previous section, Kang et al. (2007, Exp. 1) reported a different pattern. They manipulated the formats of both the intervening and final tests (i.e., multiple-choice and short-answer questions). The results of their Experiment 1 found out that taking multiple-choice tests produced superior testing effects regardless of the format of the final test, and initial short-answer tests produced better performance than a study-only (i.e., control condition) only when the final test was a short-answer format. Please note that in Experiment 2, they found that initial short answer questions produced the greater testing effect independent of the nature of the final test. This difference seems to originate from the introduction of feedback in Experiment 2. Feedback has differential effects on the testing effect. Kang et al. (2007) reported that giving feedback increased the performance of the participants only in the short answer condition whereas it did not affect the multiple choice performance. On the other hand, Butler and Roediger (2007) did not find any differential effect of feedback on the format of the tests. The feedback condition and no-feedback condition did not differ from each other on the final criterion test. Although this issue deserves further empirical research, no

further attention will be given to it here since feedback is not in the scope of the present study. One of the objectives of the present study is to replicate Kang et al.'s (2007) and Carpenter and DeLosh's (2006) study and elaborate on the experimental design. The present study manipulated the formats of both the initial and the final criterion test. The difference of this study from the previous ones is the design and the material. Instead of using a mixed factorial design (e.g., Carpenter & DeLosh, 2006; Kang et al., 2007), a within-subject design is utilized in the present study. Another difference of this study from the previous ones is that the experimental materials are not composed of word lists but of passages, which permits the evaluation of the transfer of the testing effect with materials resembling more typical class curriculum.

### **1.5.3. Transfer in the Knowledge Domain**

*Knowledge domain* is another dimension of the transfer where the individual transfers what s/he learns in one context setting to another. For example, after being taught theories of learning in a class, if the student uses this knowledge outside the classroom (such as when raising a child) he is considered to engage in far-transfer. If this student uses this knowledge in the very same class to comprehend another concept, such as long-term potentiation, he is considered to accomplish near-transfer.

Unfortunately, to our knowledge there is only one study which has investigated the transfer of the testing effect in the domain of knowledge. In other words, no other study systematically investigated whether superior learning observed with testing effects can be utilized in different problems and with different contexts. As the literature review cited above, the traditional procedure followed in the testing effect studies does not leave any room to evaluate the transfer of the testing effect in the knowledge domain. In these studies, after participants are presented with to-be-remembered material, they are given either another study

trial where they restudy the material or a test trial where they practice retrieving the material. Later on, a final memory test is given to assess the amount of learning in both conditions. The questions in the final memory tests used in the studies are identical to those presented in the initial tests. In this procedure, it is problematic to assess the effectiveness of testing effects since one cannot unarguably assert the testing effect reflects deeper levels of learning. The participants in these studies might have memorized the items (i.e., rote learning) and used the memorized information to answer the final test. This issue was brought up by Rohrer, Taylor and Sholar (2010) very recently. They argued that in most of studies, the initial and final tests included the same questions which does not allow us to assess the practical utility of the testing effect. They investigated this issue by using a visual task (e.g., regional maps). 4<sup>th</sup> and 5<sup>th</sup> grade students studied the maps. The test was either assigning the region names to the map locations, i.e., the same as in the initial test condition, or identifying the city they would pass through as they drove along the shortest possible route between two given cities, i.e., transfer task. Rohrer et al. (2010) found out that the testing effect did not decrease in the transfer task. Actually it was slightly larger in the transfer task. This study is of great value because it showed that with a visual task, the memory gain induced by testing effect can be utilized even when asked in different ways. The present study aims to take this issue further and investigate transfer of the testing effect in knowledge domain with verbal materials and with university students. Moreover, the present study also aims to assess the near and far transfer in the knowledge domain. The primary motivation for the study is the testing effect could be even more valuable if it produced learning of a complex fact (i.e., higher level learning) rather than learning of a particular answer to a particular question (i.e., rote learning) (McDaniel et al., 2007). Learning of a particular answer to a particular question is an example of rote learning which seems, as in Bloom's example cited

above, to prevent the transfer of learning. Rote learning is certainly valuable, but it is considered to be very different from “true knowledge” (p. 29) which can be utilized in new situations and in a very different form than which it was originally learned (Bloom, 1956). It is important to examine whether the testing effect produces *true knowledge* or *rote learning* for theoretical and practical reasons. The answer to this question would provide empirical grounds to differentiate theories of the testing effect which make different predictions about the mechanisms underlying the testing effect. This issue is discussed in the following section in detail. Moreover, the testing effect and its transfer in the knowledge domain bear important practical importance in classroom settings. Instructors are hesitant to ask identical questions on quizzes or midterms and on final assessment, as most of these testing effect experiments do. If the testing effect promotes extraction of information from the test, instead of verbatim copy of the learned material, then the learner can use this extracted information, even if he or she is tested in a different format or context. There is not much work investigating this issue. McDaniel et al. (2007) attempted to examine this question by changing the questions in the initial quiz and the final test in the following way.

Initial test: All preganglionic axons, whether sympathetic or parasympathetic release \_\_\_\_\_ as a neurotransmitter.

- a. acetylcholine
- b. epinephrine
- c. norepinephrine
- d. adenosine

Final test: All \_\_\_\_\_ axons, whether sympathetic or parasympathetic release acetylcholine as a neurotransmitter.

- a. preganglionic
- b. ionotropic
- c. hypothalamic
- d. adenosine

McDaniel et al. (2007) found a robust testing effect: initial testing improved memory performance on subsequent memory tests. Based on the results, they argued that the testing effect does not reflect only rote learning since they tested “retention of the complete conceptual relation” (p.500). However, the results of this study should be interpreted with some caution at least for two reasons. First of all, they did not include any *control condition* where the initial test and the final test included the same questions. To eliminate role of simple rote learning in testing effects, we need to see that students in the *control condition* (i.e., same wording in initial test and final test) and experimental condition (i.e., different wording in the initial test and the final test) demonstrate similar performance. The second problem is in the operationalization of the testing effect. McDaniel et al. (2007) argued that their design measures the testing effect as “retention of the complete conceptual relation” rather than “retention of a particular answer provided in a quiz” (p. 500). In other words, different wording between the initial test and the final test provides the means of measuring the retention of the conceptual learning but not verbatim learning of the facts. This argument also is open to discussion: the questions in the final test (as can be seen above) are mere replications of the questions in the initial test except the answer required for a question in the initial test is now embedded in the question stem in the final test.

Therefore, students might have relied on rote learning once again. Therefore, the present study aims to control this factor by rephrasing the questions in the initial and the final test.

#### **1.5.4. Evaluation of the theories with respect to the transfer of the testing effect in modality, temporal and knowledge domain**

Studying the transfer of the testing effect is important for theoretical purposes because each of the theories of the testing effect makes different predictions about the transferability of the testing effect. As noted above, one of the aims of the present study is to compare two dominant theories, the retrieval hypothesis and the transfer-appropriate processing view.

The retrieval hypothesis argues that the testing effect reflects more than rote learning. According to this view, at the time of testing, individuals engage in deeper cognitive processes which, in turn, proliferates memory traces, i.e., that is encoding variability. At the time of retrieval, an individual uses one of the memory cues to recollect information. According to this view, retrieval effort at the initial test, i.e., difficulties induced at the time of testing as opposed to rereading the materials, promotes higher order learning which generates greater performance capacities in transfer tasks (Schmidt & Bjork, 1992). In other words, a learner extracts the information in the initial testing and transfers this information into other domains. As suggested by Rohrer et al. (2010), the testing effect should not decrease in the transfer task.

The transfer-appropriate processing view makes very specific predictions. As reviewed above, this framework postulates that memory performance benefits from the processes which are captured at the time of encoding (Rajaram, Srivinas & Roediger, 1998). Therefore, for this framework, the transfer is a function the similarity between learning contexts and testing contexts: the more similarity between two contexts, the better the transfer. As the similarity between encoding and retrieval processes decrease, the testing effect should decrease. As

described in the method sections below, in the present study, different levels of resemblance were manipulated to test the predictions of this framework.

To sum up; the purpose of the present study is to examine whether the testing-induced memory gain (i.e., the testing effect) is transferable to other tasks or whether it reflects rote learning. To accomplish this objective sheds light in the theoretical arena because each of the theoretical frameworks makes a different prediction concerning the transfer of the testing effect.

## **2. The Present Study**

The transfer of the testing effect is assessed in the present study. The main research question is as follows: “How efficiently can students utilize newly learned information via testing effects in different contexts and forms?” To investigate this question, the present study aims to assess the transfer of testing effects in the temporal domain, the modality domain and the knowledge domain and to evaluate theories of the testing effect.

To accomplish the objectives of the study, three different types of manipulations were employed: for temporal domain; for knowledge domain and for modality domain. As it is described in detail in each method section, only a summary of the manipulation is provided in this section.

According to the literature review above, the testing effect lasts over an extended period of time, such as 1 week (e.g., Karpicke & Roediger. 2006a), and 1 semester (e.g., McDaniel et al., 2007). Therefore, it can be safely postulated that the transfer of testing effect is expected to occur in the temporal domain. The present study, except Experiment 3, employed a 1-week delay manipulation. In all of the experiments, except Experiment 3, the participants studied the materials, and in accordance with the condition, they were given additional time to restudy or were given an intervening test. After a week, the participants took a final test which measured

their recollection of the materials. Experiment 3 was conducted in an actual classroom setting incorporating real classroom curriculum. The retention interval varied between 4-12 weeks.

One of the most important objectives of the present study is to examine whether the testing effect is limited to the surface level of learning (i.e., rote learning) or is extended to the deeper levels of learning (i.e., *true knowledge* in Bloom's terms or *conceptual retention* in McDaniel et al.'s terms). One way of investigating this question is the examination of the transfer of testing effects in the knowledge domain which is very crucial because it gives us information about the depth of the testing effect. Nevertheless, as compared to the temporal domain and the modality domain, the transfer in the knowledge domain is the problematic field of study: there is only one study, to my knowledge, which investigated the testing effect in the knowledge domain. As a result, the predictions can be drawn based on the theories with no complementary empirical studies. Furthermore, the theoretical frameworks make different predictions concerning the transfer of the testing effect in the knowledge domain which will be discussed shortly. Three experiments in this study (Experiment 1, Experiment 2, and Experiment 3) were dedicated to evaluate the transfer, which ranges between near and far, of testing effects in knowledge domain. Near transfer occurs when questions are similar in the initial test and final tests. Nearly all of the studies on the testing effect investigated this kind of transfer. The designs of Experiment 1, Experiment 2 and Experiment 3 aim to assess near and far transfer by manipulating the similarity between learning contexts and testing contexts. In Experiment 1 and Experiment 3, the questions in the initial and the final tests have different wording. In other words, the final test contains the rephrased version of the questions of the initial tests, which constitutes the near transfer condition. The transfer appropriate processing view predict that, as the similarity between the initial test and the final test increases, the testing effect will be



stronger. Rephrasing the questions in the final test is expected to interfere with the transfer of the testing effect. On the other hand, the retrieval hypothesis predicts that the testing effect represents not a surface processing but a deeper, relational processing; therefore, retention should not be affected as much by wording similarity between the initial test and the final test. In other words, rephrasing the words should interfere with the magnitude of the testing effect in a way that memory performance should be significantly lower in the rephrased final questions according to the transfer appropriate processing view whereas, according to the retrieval hypothesis, rephrasing the questions should not tamper with the testing effect since the testing effect is a robust phenomenon which is more related to the underlying memory processes than superficial processing.

In addition to near transfer, the present study also planned to assess far transfer of the testing effect in the knowledge domain. In Experiment 2 and Experiment 3, the content of the questions were manipulated (i.e., factual or conceptual/applied). *Factual questions* refer to factual information that the students will study beforehand. For example, the following question is a factual question:

- The fact that most people tend to perceive visual elements that are closer to each other as being a single group illustrates the Gestalt principle of
  - a. similarity
  - b. continuity
  - c. **proximity**
  - d. good form

The conceptual form of this question is;

- Ray sat on his porch looking out at his field of corn. The fact that Ray perceived the corn in the field as being grouped into a series of separate rows is consistent with the Gestalt principle of
  - a. similarity
  - b. continuity
  - c. **proximity**
  - d. good form

Both of the questions ask about the same concept: proximity. For the transformation of questions, Bloom's (1956) taxonomy was used as a guide. Bloom defined six levels of cognitive domain: knowledge, comprehension, application, analysis, synthesis, and evaluation. Knowledge level represents the lowest level of learning where a learner memorizes, defines or duplicate the learned material. In each ascending level, the learning processes require more complex and abstract mental levels. Bloom (1956) argued that the majority of the test questions students are exposed force them to think at the lowest level, namely knowledge level. In Experiment 2 and Experiment 3, knowledge level questions, which are labeled as *factual questions* in the present study, were transformed in to application level questions, which are labeled as *conceptual questions*. Application level questions aim to force the students to use the acquired knowledge in different ways to solve the new questions. The conceptual questions are constructed solely for this purpose: to compel the participants to use the memory gain in different contexts which requires comprehension of the material rather than memorizing.

In Experiment 2 and Experiment 3, the nature of the questions (i.e., factual or conceptual) was manipulated factorially. For example, when a participant took a factual initial test, sometimes s/he took a final factual test and sometimes took a conceptual version of a final test.

This design allowed for assessment of far transfer of the testing effect because the context and the form of encoding (i.e. factual) is very different from the context and form of the final test (i.e. conceptual). Furthermore, to be able to answer the conceptual versions, an individual has to engage in higher level encoding processes and in deployment of factual information. If the testing effect really promotes relational processes, an individual should be able to transfer between factual and conceptual problems. According to the retrieval hypothesis, the testing effect should be transferable between conceptual and factual domains since testing promotes deeper and relational processing which allows for the extraction of the gist of the problem. The transfer-appropriate processing framework predicts that as long as the processes during encoding and retrieval match, the transfer should be observed. Therefore, this framework predicts that the conceptual initial test should produce the greater testing effect when the final test is conceptual than when the final test is factual. Similarly, the factual initial test should produce the greater testing effect when the final test is factual than when the final test is conceptual. On the contrary, the conceptual initial test should produce an insignificant testing effect when the final test is factual and vice versa.

Experiment 4 aims to examine the transfer of the testing effect in the modality domain. The research on the transfer of the testing effect on modality has yielded contradictory results. In addition to that, the theories of the testing effect make different predictions. Therefore, it is not easy to make straightforward predictions about the transfer of testing effects across different modalities. In the present study, modality is operationalized as the format of the initial test and the final test which can be a recall test or a recognition test. The retrieval hypothesis predicts that taking the initial *recall* test produces the best testing effect and the best transfer of testing effects, regardless of the format of the final test. This is because, according to this view, retrieval effort

during the encoding phase determines how well the information is retained. Traditionally, recall tests are accepted as more difficult than recognition tests (for a review see Gillund & Shiffrin, 1984) because recall requires more extensive recollection of the to-be-learned information (Haist, Shimamura & Squire, 1992). On the other hand, according to the transfer-appropriate processing view, transfer should be the best when the initial and the final test formats are the same. In Experiment 4, the formats of the initial and final test were manipulated. The participants took either a recognition task or a recall task as an initial test. After a week, they took the final test either in the same format as the initial test or in a different format than the initial test. This design allows us to test the predictions of the retrieval hypothesis and the transfer appropriate processing view.

## **2.1. Experiment 1**

In Experiment 1, I manipulated the wording of the questions on the initial and final test in a way that the questions in the initial and final test had the same wording or different wording (but answers were always the same). Wording of the questions makes a profound difference in assessing learning as illustrated with the example by Bloom's "Igneous fusion" example (1956) above. In addition, this experiment also explores a certain aspect of ecological validity for the testing effect: Since few professors are willing to ask identical questions on midterms and finals, as the vast majority of testing effect experiments do, the present study examined if benefits of testing are maintained even with different wording. It is hypothesized that if testing effects promote deeper levels of learning, then participants will be able to use the memory gain in different forms (i.e., different wording). On the other hand, if the testing effect is a result of rote (exact) memorization, then participants will not be able to utilize the memory gain in a different form (i.e. different wording).

### **2.1.1. Method**

#### **2.1.1.1. Participants**

Participants were 47 undergraduates (age range = 18-25 years) from the introductory psychology subject pool at Rice University. However, the data from 39 participants were used in the statistical analysis due to the selection criterion of the experiment. Please see the materials section for detailed information. All participants received course credit in exchange for volunteering.

#### **2.1.1.2. Materials**

*Study passages.* Three essays were created for the purposes of the study. Each essay covered a certain topic in psychology (i.e., Attribution Theory, Gestalt Psychology, and Personality Theory). There were no tables or figures in the essays. The length of the essays ranged between 712-770 words (See Appendix A).

*Tests.* From each essay, 10 facts were selected. These facts were tested in multiple-choice format (i.e., recognition task) in which participants had to choose a response among four options. For each essay, two sets of multiple-choice questions were created: Form A and Form B, both of which asked about the same facts (See Appendix B). The only difference between the forms was the phrasing of the questions. For example, one particular question in Form A was as follows:

People tend to make external attributions for a behavior when that behavior is determined by

- a. personal dispositions, traits, abilities and feelings
- b. factors that the individual has no control over

c. **situational demands or environmental factors**

d. factors that the individual is able to control

The re-worded version of the previous question in Form B was as follows:

An external attribution occurs when individuals attribute the cause of another person's behavior to

a. personal dispositions, traits, abilities and feelings

b. factors that the individual has no control over

c. **situational demands or environmental factors**

d. factors that the individual is able to control

Note that the answer options and the answers, in bold, of the questions were still the same. To control the difficulty level of Form A and Form B, a pilot study was conducted. We found that the average score obtained in Form A and Form B did not differ significantly ( $p > .10$ ). The participants were also given a demographic questionnaire which aimed to assess their knowledge of the topics covered by the essays. In the demographic questionnaire (See Appendix C), the participants were asked whether they are familiar with the concepts that they have read in the experiment. If the participant had any familiarity, s/he was asked to rate his/her level of knowledge on a 5-point Likert scale (1=Poor to 5=Expert). The participants with a score 3 or above were not included in the data analysis due to the possibility that their existing knowledge of the concepts could interfere with the manipulation of the experiment.

#### 2.1.1.3. Design

A 2 (initial encoding task: multiple-choice test or reread the essays) X 3 (final test questions: new questions, old questions or re-phrased questions) within subject design was used. During the first session, the participants studied the three essays. In accordance with which group they were assigned to, immediately after reading the first essay, the participants either took a multiple-choice test or were given more time to reread the essay. In the second session, which was held after 1 week, the participants were tested on all of the essays. In the final test, they were given a multiple-choice test which included three types of questions: new questions, old questions and rephrased questions. New questions were not asked in the initial test and were from the essays that the participants were asked to reread in the first session. Old questions were asked in the initial test. Rephrased questions were asked in the initial test but they were reworded in the final test. The order of the essays and the initial encoding task were fully counterbalanced across subjects.

#### **2.1.1.4. Procedure**

Participants were tested in groups of 10 or fewer during two experimental sessions. During the first session, the participants were seated and given paper copies of the essays one at a time. At the beginning of the experimental session, participants were told that their memory of the essays would be tested. They were told not to underline or mark any part of the essays. They were given 10 minutes to read the each essay. After the 10 minutes elapsed, the participants were instructed to put away the paper copy of the essay. In the test condition, the participants were given 10 minutes to complete a multiple-choice test, whereas in the restudy condition, the participants were given 10 more minutes to reread the essay. Once they were done with the task (i.e., multiple-choice test or rereading the essay), the participant proceeded to read the next essay. The first session lasted about 1 hour.

One week later, the participants returned for the second session. They were tested on all three essays that they had read in the first sessions. They were again tested in a multiple-choice format. After they returned the completed multiple-choice test, the participants were given a demographic questionnaire which assessed their knowledge of the topics which were covered in the essays. After they filled out the demographic questionnaire, they were debriefed and dismissed. The second session lasted 15 minutes.

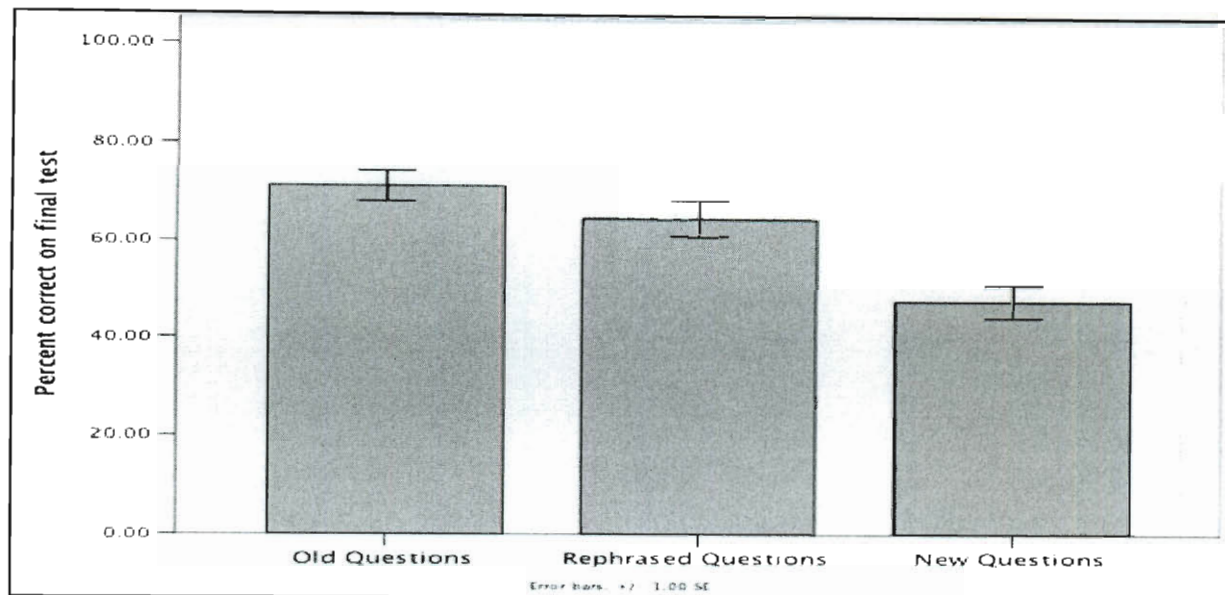
### 2.1.2. Results

Responses were counted as either correct (1 point) or incorrect (0 point). It was found that the final performance was affected by the initial task: taking a test on the material improved memory performance on the final test compared to the restudy condition,  $t(37) = 5.78, p < .05, d = .97$ .

To assess the depth of the testing effect, the percentage of questions answered correctly by each participant was computed separately for the final test questions: old questions ( $M = 71.31, SD = 19.19$ ), new questions ( $M = 47.63, SD = 20.85$ ) and rephrased questions ( $M = 64.73, SD = 22.74$ ). There was a significant main effect of the type of questions,  $F(2, 74) = 23.92, p < .05$ . Post hoc comparisons using paired-samples t-tests indicated that new questions yielded poorer performance than performance on the old questions and rephrased questions,  $t(37) = -6.36, p < .05, d = 1.18$ , and  $t(37) = -4.39, p < .05, d = .78$ , respectively. The participants displayed higher performance with old questions than rephrased questions,  $t(37) = 2.26, p < .05, d = .31$  (See Figure 2). The results showed that participants' recognition memory performance was superior when the questions on the final test were the same as in the initial test. On the other hand, participants demonstrated poor recognition performance with new questions on the final



test.



**Figure 2. Mean percentage of correct answers as a function of the type of questions asked on the final test in the laboratory study of Experiment 1.**

## Discussion

Experiment 1 was conducted to assess the transfer of the testing effect in the temporal domain and in the domain of knowledge. The results of Experiment 1 showed that performance on the final test improved when participants took an initial test. Even information tested in a different format after 1 week – with different wording on a final test compared to an initial test – still showed a benefit of previous testing. In other words, when an individual utilizes tests as learning aids, s/he can still use the memory gain produced by tests on later tests. This finding has important practical implications as noted above. Giving frequent quizzes might help the students for a later exam, even with the rephrased questions. It seems that students are not only tested about their knowledge in the test but also they learn from the test. Moreover, encoded

information does not seem to be of a superficial level. They can still access the encoded information even when they have been asked in a rephrased format after 1 week delay.

Although the results of the experiment provided evidence for the transfer of the testing effect in the knowledge domain and in the temporal domain, it raises an important concern: depth of the transfer. As reviewed above, there are different dimensions of the transfer changing on a continuum of near and far transfer. According to Barnett and Ceci (2002), the similarity between learning and testing conditions determines the depth of the transfer. In Experiment 1, simply rephrasing the questions does not enforce participants to engage in higher level cognitive processes. Hence, inarguably, the depth of the transfer is limited to near transfer. Experiment 2 was conducted to tackle this concern.

Another important concern of Experiment 1 raises is the role of rote learning in the testing effect. The results showed that although the participants can use the test-induced memory gain with rephrased questions, their performance is still superior when they are tested with the same questions (i.e., old questions), which indicates that the testing effect involves some amount of rote learning. This finding has important bearing on questions about the theoretical orientations explaining the testing effect and will be discussed under general discussion below.

## **2.2. Experiment 2**

Experiment 2 assessed the transfer of the testing effect in the temporal domain and far transfer of the testing effect in the knowledge domain. As indicated above, the transfer is a function the similarity between learning context and testing context. As the similarity diminishes, the transfer is considered to be far transfer which was accomplished by manipulating the context of the questions (i.e., factual or conceptual/applied). Factual questions ask for factual information that the participants have studied beforehand. On the other hand, conceptual questions ask for

qualitative answers and require more than simply recalling or recognizing the answer and are restricted to the level of comprehension and applications in Bloom's terms (1956). In this regard, the conceptual questions will not be the verbatim repetition of the facts stated in the studied texts. Instead, as noted above, conceptual questions require the participants to interpret the concepts in the texts and use them in both abstract and concrete situations. The main question investigated in Experiment 2 is whether the participants can transfer what they learn in the factual test to a conceptual test and vice versa. If they can accomplish such transfer, then it can be argued that the testing effect is not limited to mere memorization of the to-be-tested information. Instead, the testing effect promotes relational processes where an individual engages in deeper levels of cognitive processes during tests, which is predicted by the retrieval hypothesis. In this regard, Experiment 2 was conducted (1) to evaluate whether the transfer of the testing effect extends to far transfer, and (2) with regards to the far transfer, to evaluate the predictions of different theoretical orientations explaining the testing effect.

### **2.2.1. Method**

#### **2.2.1.1 Participants**

Participants were 63 undergraduates (age range = 17-23 years) from the introductory psychology subject pool at Rice University. However, the data from 55 participants were analyzed because 6 participants indicated on the demographic questionnaire that their level of knowledge of the concepts and theories in the study passages were satisfying. Please see the Materials section below for more information about the selection criteria. The data from 2 participants were not used in the analysis due to failure to follow the experimental instruction. All participants received course credit in exchange for volunteering.

### 2.2.1.2. Materials

*Study passages.* Five essays were created for the purposes of the study. Each essay covered a certain topic in psychology (i.e., Gestalt Psychology, Attribution Theory, Seven Sins of Memory, Freud's Personality Theory and Personality Theory). There were no tables or figures in the essays. The average length of the essays was 782 words ranging between 710 and 930 words (See Appendix D).

*Tests.* From each essay, 5 facts were selected. These facts were tested in multiple-choice format in which participants had to choose a response among four options. For each essay, two sets of 5 multiple-choice questions were created based on these 5 facts (i.e., 5 Conceptual Questions and 5 Factual Questions). Both of the forms asked about the same fact (See Appendix E). The only difference was that the conceptual questionnaire consisted of conceptualized version of the factual questions. For example, one of the factual questions is as follows:

When \_\_\_\_\_ is made, the cause of the given behavior is assigned to the individual's personality, attitudes, character or disposition.

- a. an external attribution
- b. a self-serving attribution
- c. an attributional error
- d. an internal attribution**

The conceptual version of this question is:

Bella watches Jose slip on the stairs of their apartment building. If Bella decides that Jose is an uncoordinated person who was not paying attention to what he was doing, she has made

- a. an external attribution

- b. a self-serving attribution
- c. an attributional error
- d. an internal attribution**

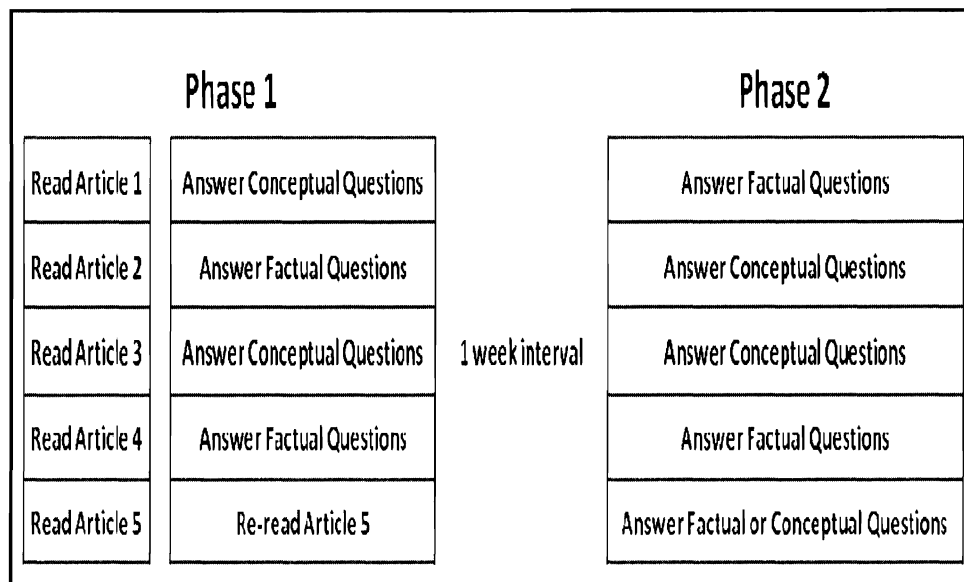
Note that the answer options of the questions were still the same.

*Demographic questionnaire.* As in Experiment 1, the participants were also given a demographic questionnaire which aimed to assess their knowledge of the topics covered by the essays. In the demographic questionnaire, the participants were asked whether they are familiar with the concepts that they have read in the experiment. If the participant had any familiarity, s/he was asked to rate his/her level of knowledge on a 5-point Likert scale (1=Poor to 5=Expert). The participants with a score of 3, which indicates satisfying knowledge, or above were not included in the data analysis due to the possibility that their existing knowledge of the concepts could interfere with the manipulation of the experiment.

### **2.2.1.3. Design**

A 3 (initial encoding task: factual questions, conceptual questions or reread the essays) X 3 (final test questions: new questions, factual questions or conceptual questions) within subject design was used. During the first session, the participants studied the five essays. In accordance with which group they were assigned to, immediately after reading the first essay, the participants either took a multiple-choice test which consisted of factual or conceptual questions or were given more time to reread the essay. In the second session, which was held after 1 week, the participants were tested on all of the essays. In the final test, they were given a multiple-choice test which included two types of questions: new questions and old (See Figure 3 for the

design of Experiment 2). New questions were not asked in the initial test and were from the essays that the participants were asked to read in the first session. There were two types of new questions: factual and conceptual. Similarly there were two types of old questions: transformed or non-transformed. Transformed old questions are the ones which are transformed to factual or to conceptual ones in the final test in accordance with the initial test. For example, factual questions in the initial test were transformed into conceptual questions in the final test and vice versa. On the other hand, non-transformed questions in the final test are the duplication of the questions in the initial test. For example, a factual non-transformed question in the final test was asked in the initial test in the same way. The order of the essays and the initial encoding task were fully counterbalanced across subjects.



**Figure 3. The design of Experiment 2.**

#### **2.2.1.4. Procedure**

Participants were tested in groups of 5 or fewer during two experimental sessions. During the first session, the participants were seated and given paper copies of the essays one at a time.

At the beginning of the experimental session, participants were told that their memory of the essays would be tested. They were told not to underline or mark any part of the essays. They were given 10 minutes to read the each essay. After the 10 minutes elapsed, the participants were instructed to put away the paper copy of the essay. In the test condition, the participants were given 5 minutes to complete a multiple-choice test, whereas in the restudy condition, the participants were given 5 more minutes to reread the essay. Once they were done with the task (i.e., multiple-choice test or rereading the essay), the participant proceeded to read the next essay. The first session lasted about 1 hour and 15 minutes.

One week later, the participants returned for the second session. They were tested on all five essays that they had studied in the first sessions. They were again tested in a multiple-choice format. After they returned the completed multiple-choice test, the participants were given a demographic questionnaire which assessed their knowledge of the topics which were covered in the essays. After they filled out the demographic questionnaire, they were debriefed and dismissed. The second session lasted 20 minutes.

### **2.2.2. Results**

Responses were counted as either correct (1 point) or incorrect (0 point). To assess the transfer of the testing effect, the percentage of conceptual and factual questions answered correctly by each participant was computed separately for the initial test (T1) and final test (T2).

#### **2.2.2.1. Initial Test Performance**

Although the focus of the study is on the final test, the initial test performance was also examined. For each participant, I computed the percentage of items answered correctly on the initial conceptual ( $M = 73.18$ ,  $SD = 23.04$ ) and factual ( $M = 84.91$ ,  $SD = 18.12$ ) tests. The

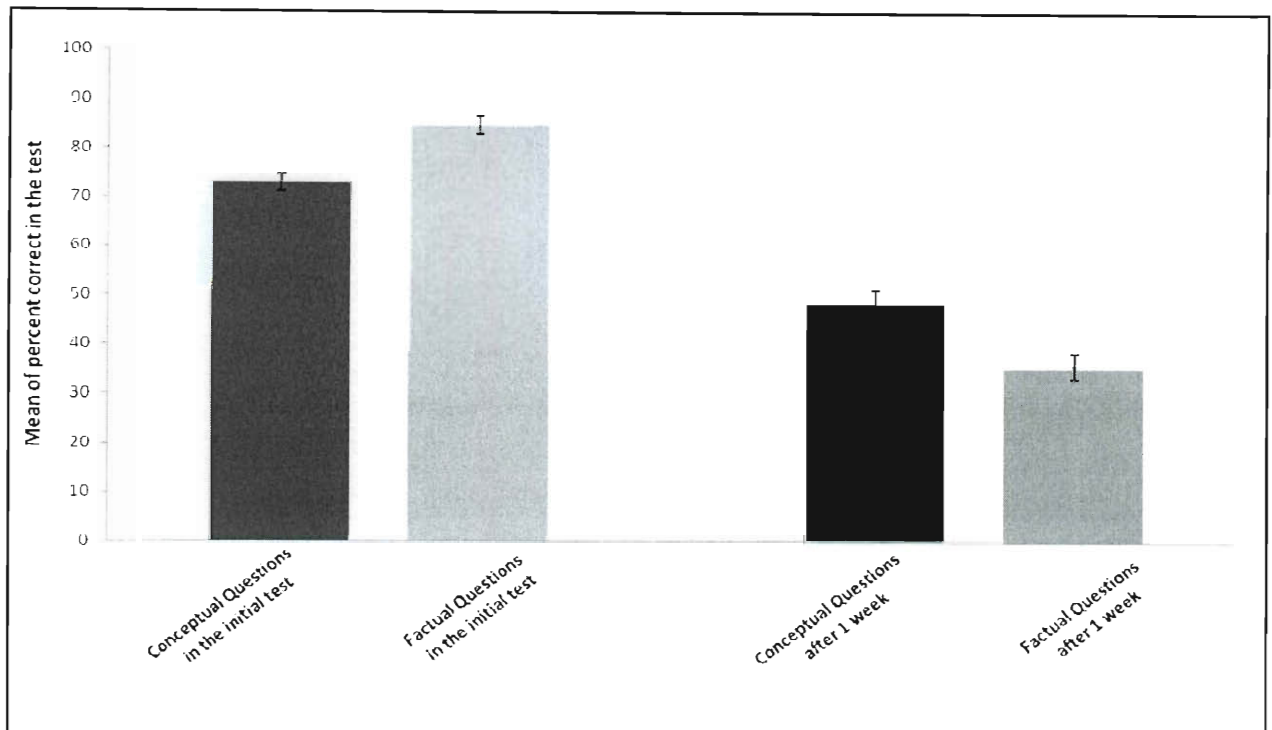
difference between two types of test was significant,  $t(54) = 5.09, p < .001, d = .67$ . The higher performance on the factual questions relative to the conceptual questions in the initial test has important implications for both their effect on the final test and for theoretical explanations, as will be discussed below in the discussion section.

#### **2.2.2.2. Final Test Performance**

The first analysis was conducted to measure whether taking a test as an initial encoding task improved memory performance on the final test (i.e., the testing effect). It was found that the final performance was affected by the initial task: taking a test on the material improved memory performance on the final test compared to the restudy condition,  $t(49) = 8.57, p < .001, d = 1.21$ .

Further analysis was conducted to examine whether the performance on the new questions (i.e., study-only condition) changes as a function of the type of the questions (i.e., factual and conceptual). It was found that the participants were less likely to remember factual information ( $M = 35.56, SD = 19.48$ ) than do they conceptual information ( $M = 48.15, SD = 22.37$ ) after 1 week delay,  $t(26) = 1.99, p = .02, d = .38$ . This finding is very important considering the fact that the recollection of the factual information in the immediate test was superior to conceptual recollection. The implication of this interaction will be discussed in the general discussion section (See Figure 4).

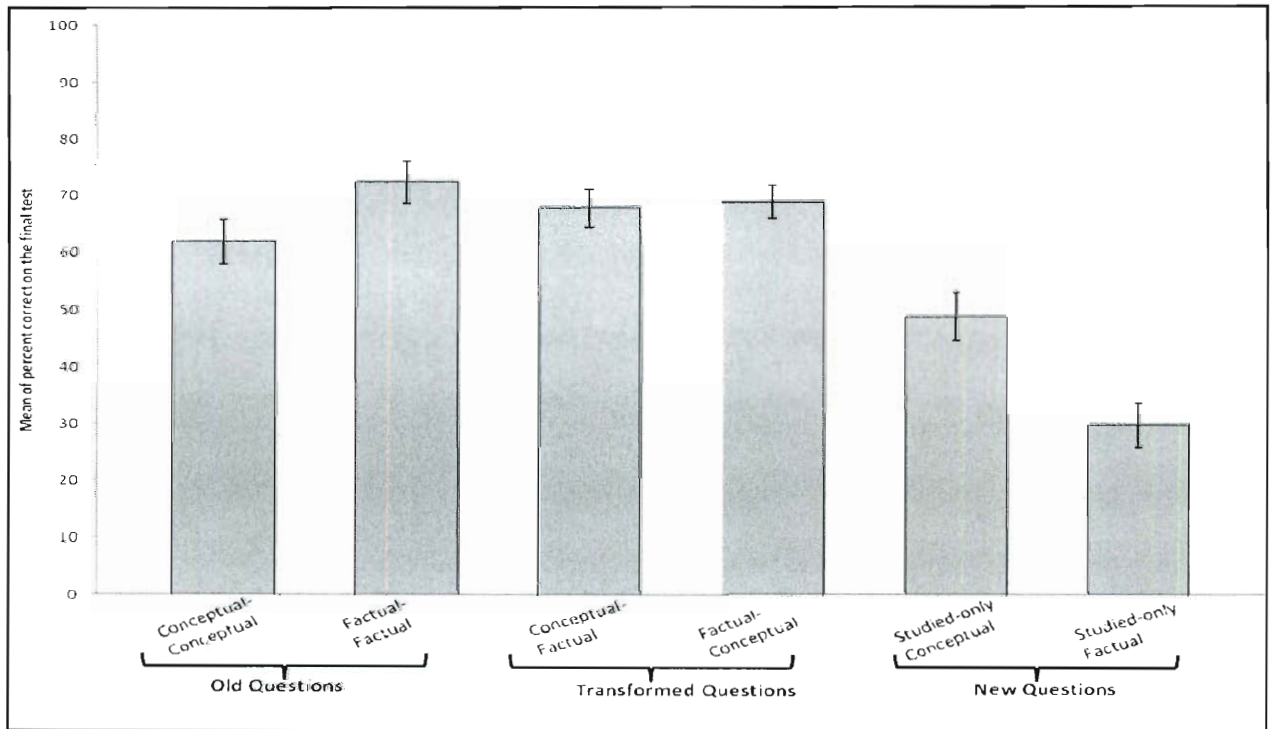




**Figure 4. The interaction between the type of questions (i.e., factual vs. conceptual) and 1 week interval.**

Since the primary interest of Experiment 2 is the transfer of the testing effect, the focus of analysis in this section is on the effect of transformation on the final test performance. If the testing effect relies on deeper levels of cognitive processes, then test-induced memory gain would be expected even though there is a mismatch between the initial testing condition and the final testing condition. The participants should be able to utilize what they learn in the factual initial test in the final conceptual test and vice versa. For the analysis purposes, four different dependent variables were created as a function of transformation: (1) Conceptual-Conceptual (CC) score, (2) Factual-Factual (FF) score, (3) Conceptual-Factual (CF) score, and (4) Factual-Conceptual (FC) score. The first word in the score label refers to the type of the initial test, and the second label indicates the type of the final test. For example, Conceptual-Factual score means that the participant took a conceptual questionnaire as an initial test and after a week, s/he took

the factual version of the test as final test. The pair wise comparisons were calculated conducted between CC and FC condition and between FF and CF. None of the pair wise comparison was significant,  $p > .10$  (See Figure 5). This finding shows that the participants could use the memory gain produced by the initial test even when they are tested in different format. In other words, the memory gain produced by testing effect is transferable into different context.

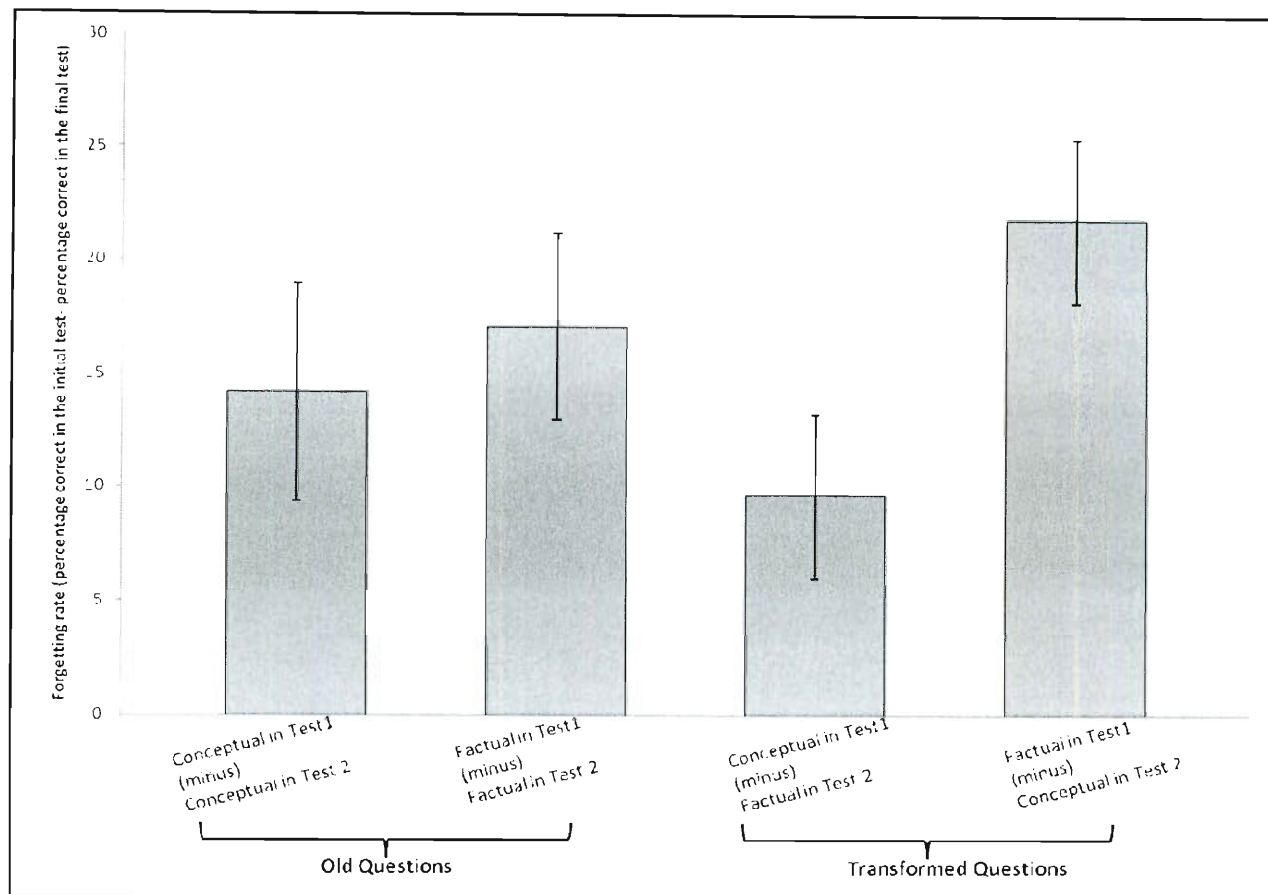


**Figure 5. The transfer of the testing effect as a function of question type (i.e., factual vs. conceptual) and the encoding type.**

### 2.2.2.3. Forgetting rate and the testing effect

Another way of assessing the transferability of the testing effect is assessment of the forgetting rate throughout the week as a function of transformation. The rationale for this analysis is to find out whether the participants' forgetting rate changes as a function of the experimental condition (i.e., transformed or non-transformed questions). For this purpose, a

difference score was obtained in a way that the percentages of correctly recognized items in the final test was deducted from the percentages of correctly recognized items in the initial test (i.e., Initial test - Final test). Therefore, four difference scores were obtained. Conceptual1 – Conceptual 2 (C1-C2) score was computed as a difference score between memory score in the initial conceptual test minus memory score in the final conceptual test. Following the same procedure, F1-F2, C1-F2 and F1-C2 difference scores were also calculated. Repeated measures of ANOVA revealed a significant difference between forgetting rates,  $F(3, 54) = 3.16, p = .02, MSE = 446.61$  (See Figure 6). Post hoc pair wise comparisons showed that the participants are more prone to forget information when they took a factual initial questionnaire and then took a conceptual questionnaire (in F1-C2 condition) after 1 week than in C1-C2 condition ( $t(54) = 1.68, p = .04, d = .23$ , and than in C1-F2 condition ( $t(54) = 2.85, p = .003, d = .38$ ). There was no difference between F1-C2 and F1-F2 condition,  $p = .12$ . The C1-F2 condition resulted in less forgetting than F1-F2 condition,  $t(54) = 1.94, p = .02, d = .26$ . The overall results revealed that factual initial test resulted in poorest transfer when the final test is conceptual. On the other hand, conceptual initial test resulted in superior transfer when the final test is factual. This finding has important theoretical and practical implication as will be discussed in the following discussion section.



**Figure 6. Forgetting rate as a function of the question type (i.e., factual vs. conceptual) throughout 1 week delay.**

### 2.2.3. Discussion

The results displayed in Figure 5 showed that taking an initial test increased final test performance as compared to simply rereading the study texts, regardless of whether the initial test included conceptual or factual questions, resulting in a robust testing effect. The results also provided evidence for the transfer of the testing effect: the participants' recognition in the final test did not change as function of transformation of the questions. Even when there is a mismatch between types of questions in the initial and final test, the participants' performance was similar to the performance in the absence of such mismatch. For example, an individual's memory

performance in the final conceptual test was comparable when s/he took a conceptual or factual initial test. Same finding was obtained for factual questions. This finding is in line with the prediction of retrieval hypothesis and it contradicts the prediction of transfer-appropriate processing view which asserts that the similarity between processes of encoding and retrieval determines the memory performance of an individual. Hence, this view predicts that when there is a mismatch between two test formats in the encoding session and retrieval session (e.g., conceptual-factual and factual-conceptual), the memory gain should be poorer than when there is a match between encoding and retrieval session (e.g., factual-factual and conceptual-conceptual) which could not be found in Experiment 2. Instead, the findings of the experiment provides strong evident for the retrieval hypothesis. The testing effect seems to reflect more than rote learning. It involves some relational processes where an individual engages in deep cognitive processes and s/he can transfer the effect into different context.

Another finding supporting the predictions of the retrieval hypothesis is the rate of forgetting which was smaller for the conceptual items. One of the assertions of retrieval hypothesis states that the item difficulty of retrieval during practice determines how well the items will be retrieved on later tests. The more difficult items, the more mental effort put on the items. The increased difficulty, and hence increased elaboration of items, creates more mental cues for later retrieval. In this experiment, conceptual questions are more difficult than the factual questions, which is demonstrated by the superior memory performance on the factual test in the initial test as compared to conceptual test. Mental processes employed in the conceptual questions apparently make these items more enduring and resistant to forgetting. The effect of item difficulty on item retrieval also explains the interaction between 1 week retrieval interval and the types of questions (e.g., factual and conceptual). In the immediate test, the participants'

performance on the factual test was superior to the conceptual test whereas after 1 week, the participants' performance on new conceptual questions was superior to new factual questions. This pattern suggests that factual information is more prone to decay than the conceptual information and the conceptual information has stronger memory traces which endure throughout a week.

## **2.3. Experiment 3**

Experiment 1 and Experiment 2 examined the near and far transfer of the testing effect in the knowledge domain and in the temporal domain in a laboratory setting. The results of Experiment 1 showed that the testing effect is transferable, albeit the effect was limited to the near transfer. On the other hand, Experiment 2 postulated that the transfer of the testing effect extends to far transfer. Since both of the experiments are laboratory experiments, one can question the ecological validity of the findings. As argued by Brewer (2000), the effect found in a variety of settings and people does not necessitate the very same effect is also observed in everyday life. This essence of ecological validity is the demonstration of the effect, which is observed in the laboratory, occurring among the population at large in their natural setting (Brewer, 2000). The question follows, then, whether the testing effect can be obtained in an actual classroom setting. As noted above, very few studies systematically studied the testing effect in the classroom (e.g., McDaniel et al., 2007). In this regard, the aim of Experiment 3 is to demonstrate the occurrence of the testing effect and the transfer of the testing effect in an actual classroom setting incorporating actual classroom curriculum.

### **2.3.1. Method**

#### **2.3.1.1. Participants**

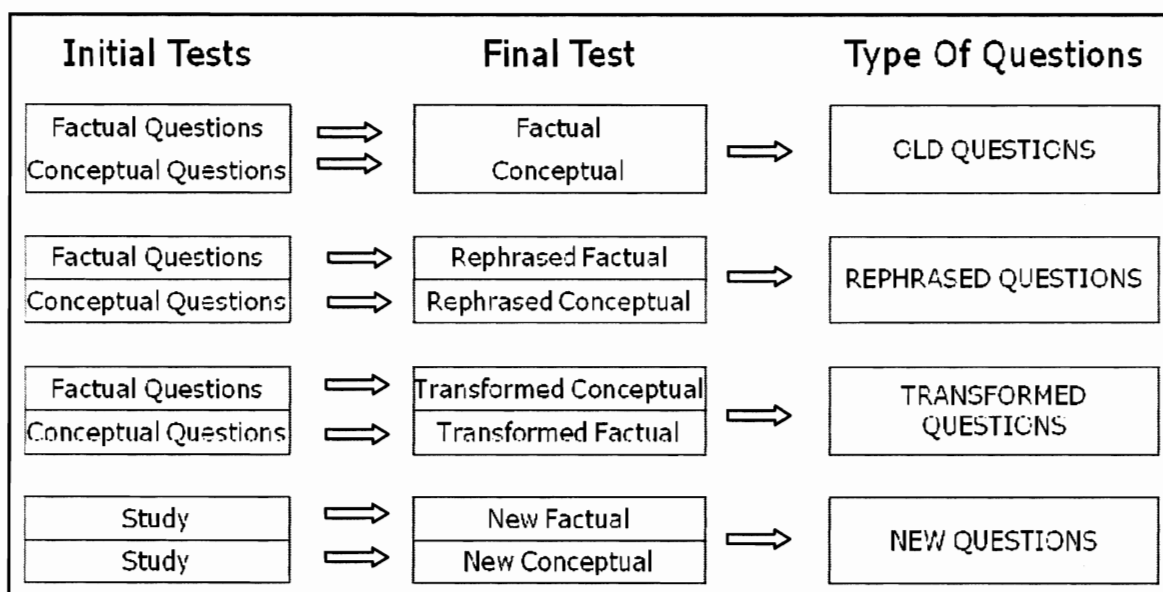
Participants were 46 undergraduates from an Introduction to Psychology class at Rice University. All participants completed three exams throughout the semester (i.e., 4 weeks apart) and a final exam which was constructed for the experiment and did not affect the grades of the students. All participants received 1 course credit in exchange for volunteering to take the “practice” (experimental) final exam.

### **2.3.1.2. Materials**

The students took three multiple-choice exams throughout the semester as a requirement of the class. Each exam covered 4 chapters from the introductory book and none of the exams were cumulative. The exams were 4 weeks apart. These exams were not a part of the experiment.

For the purposes of the experiment, a final exam composed of 93 multiple-choice questions with four answer options was constructed which included old questions, new questions, rephrased questions and transformed questions which are explained below (Also see Appendix F). To construct the final exam, a total of 70 questions from three exams were selected. Out of 70 questions, 24 of them were from Exam 1, 24 were from Exam 2 and 22 were from Exam 3. Half of the questions ( $n=35$ ) were factual questions whereas the other half ( $n=35$ ) was conceptual questions. Approximately one third of the randomly selected factual and conceptual questions ( $n=25$ ) were transformed into conceptual and factual questions, respectively and vice versa. These questions (called *transformed questions* from now on) compose one group of questions in the final test. The next group of questions ( $n=25$ ), which is called *rephrased questions*, was generated by simply rewording the questions. For example, a factual question of “Random assignment ensures that ...” is rephrased to “The main function of random assignment is ...” The same procedure was employed to rephrase conceptual questions.

The third group of questions ( $n=20$ ) in the final exam was called *old questions* which were the factual and the conceptual questions to which the students were exposed in the initial tests. The last group of questions is called *new questions* ( $n=23$ ) which were not asked in the initial tests yet the students studied the material for their exams. These questions are also called *studied-only items* (See Figure 7 for the construction of the final exam). A within subjects design was utilized. Each student answered all types of questions. Please note that none of the questions or related facts was asked on the actual in-class final exam.



**Figure 7. The construction of questions of the final “practice” test.**

### 2.3.1.3. Design

A 3 (initial encoding task: factual question, conceptual questions or study-only items) X 4 (final test questions: transformed questions, rephrased questions, old questions and new questions) within subject design was used.

### 2.3.1.4. Procedure

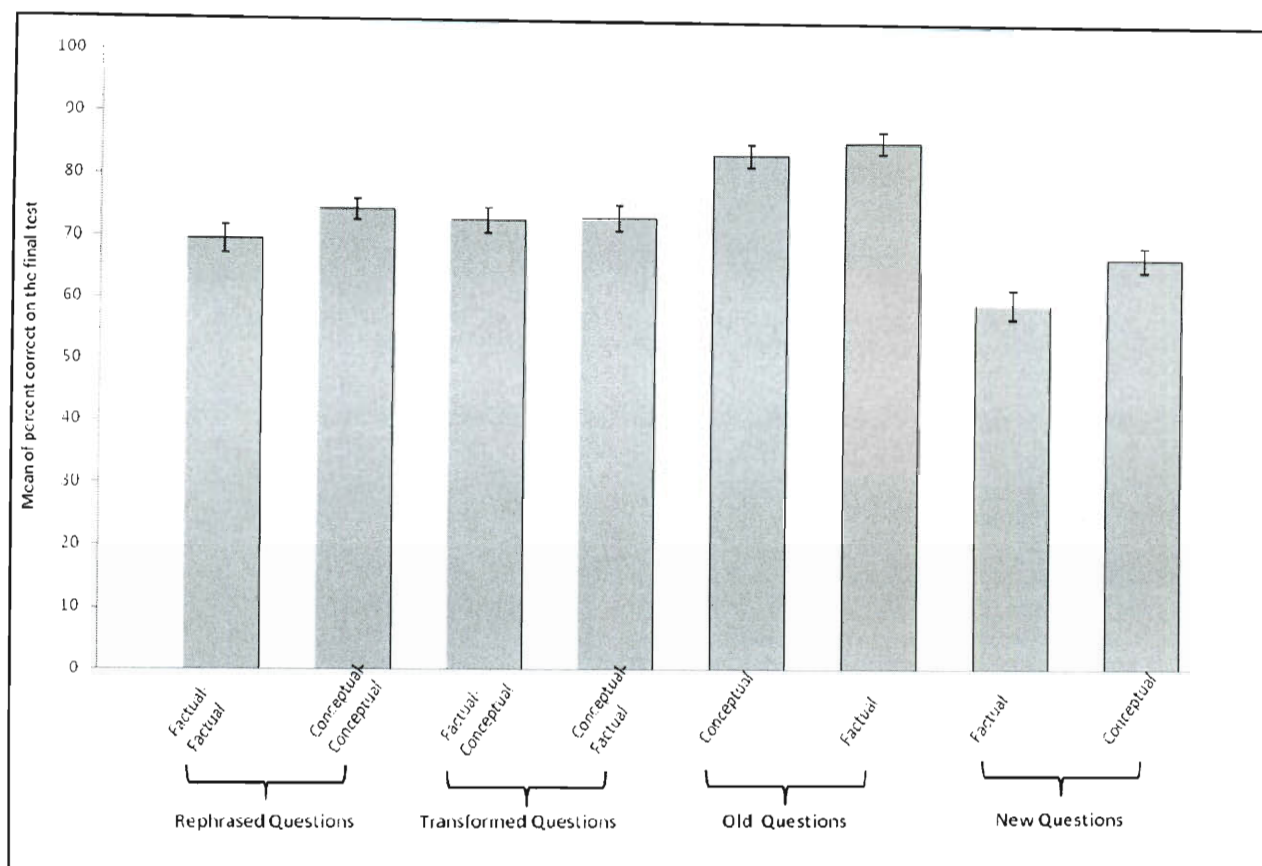


As a class requirement and independent of the experimental purposes, students took three exams throughout the semester and one final exam at the end of the semester. Three weeks after the third exam, they were told that they could take a practice final exam which might help them on the in-class final exam. The experimental session took approximately 60 minutes.

### 2.3.2. Results

Responses were counted as either correct (1 point) or incorrect (0 point). It was found that the final performance was affected by the initial task: taking a test on the material improved memory performance on the final test compared to the study-only condition,  $t(45) = 8.67$ ,  $p < .05$ ,  $d = 1.19$ . Students' performance on the previously tested questions ( $M = 78.53$ ,  $SD = 8.43$ ) was superior than previously studied items ( $M = 63.25$ ,  $SD = 11.93$ ).

To assess the depth of the testing effect, the percentage of correct answers was computed separately for the final test questions: old questions ( $M = 84.56$ ,  $SD = 9.99$ ), new questions ( $M = 63.24$ ,  $SD = 11.93$ ), rephrased questions ( $M = 71.97$ ,  $SD = 10.89$ ) and transformed questions ( $M = 73.01$ ,  $SD = 11.39$ ). There was a significant main effect of the type of questions,  $F(3, 45) = 44.61$ ,  $p < .05$ ,  $MSE = 77.81$ . Pairwise comparisons showed that the students' performance on new questions was poorer than on old questions, rephrased questions, and transformed questions,  $t(45) = -10.46$ ,  $d = -1.54$ ,  $p < .05$ ,  $t(45) = -3.99$ ,  $d = -.58$ ,  $p < .05$ , and  $t(45) = -4.45$ ,  $p < .05$ ,  $d = -.66$ , respectively, indicating that taking a test improves memory for later retrieval. Students' accuracy with old questions were superior to new questions,  $t(45) = 10.46$ ,  $p < .05$ ,  $d = .61$ , to rephrased questions,  $t(45) = 7.59$ ,  $p < .05$ ,  $d = 1.12$  and to transformed questions,  $t(45) = 7.46$ ,  $p < .05$ ,  $d = 1.11$ . There was no significant difference between rephrased questions and transformed questions. The overall pattern demonstrated that new questions led the poorest performance whereas old questions yielded the best performance (See Figure 8).



**Figure 8. The transfer of the testing effect as a function of question type and the type of encoding task.**

Further analyses were conducted to evaluate the transfer of the testing effect as a function of the type of the questions (i.e., factual vs. conceptual) on the final test. Percentage correct of conceptual questions ( $M = 74.93$ ,  $SD = 7.49$ ) was significantly higher than of factual questions ( $M = 71.47$ ,  $SD = 9.83$ ),  $t(45) = 2.65$ ,  $p = .01$ ,  $d = 0.39$ . This finding indicates that students are able to recollect conceptual information better than factual information. To obtain a detailed picture of the pattern, further analysis was conducted to evaluate the transfer of the testing effect as a function of question type (i.e., factual vs. conceptual) and type of encoding

task (i.e., transformed questions, rephrased questions, old questions and new questions). For analysis purposes, 8 variables were created: Factual-Factual, Factual-Conceptual, Conceptual-Conceptual, Conceptual-Factual, Old Conceptual, Old Factual, New Conceptual, and New Factual. The first word in the label refers to the initial test whereas the second one refers to the final test. For example, Conceptual-Factual question means that this item was asked in the conceptual way in the initial test and factual way in the final test and so on. Following the same logic, these labels can be categorized in the following way. Factual-Factual and Conceptual-Conceptual questions are the rephrased questions whereas Factual-Conceptual and Conceptual-Factual questions are the transformed questions. As noted above, old factual and conceptual questions are the replicas of the questions which were asked in the previous exams. New factual and conceptual questions are the ones which were not asked in the previous exams but the materials were covered in the lectures and the students were required to study them for the exams. Repeated measures of analysis revealed a significant difference between different question types,  $F(45,7) = 23.78, p < .001, MSE = 136.40$  (See Figure 8). Supplementary pair wise analyses showed that Conceptual-Conceptual rephrasing resulted in enhanced recollection than did Factual-Factual rephrasing, which supports the argument that conceptual information is more resistant to forgetting. There was no significant difference between two types of transformed questions (i.e., Factual-Conceptual and Conceptual-Factual),  $p < .10$ . The memory performance on the old questions did not differ as a function of factual and conceptual questions,  $p < .10$ . New questions were affected by the type of question: students scored higher on new conceptual questions than new factual questions,  $t(45) = 2.81, p = .007, d = .42$ . This finding provides another supporting evidence for the influence of conceptual processing on long term recollection as will be discussed below.

To results of Experiment 3 demonstrated the transfer of the testing effect: Even when the questions in initial test were rephrased or transformed in the final test, student's performance was superior to studied-only items, which indicates students can transfer what they learn in the exams into different test conditions. However, the apparent superiority of old questions demonstrates that rote memorization is still evident in the testing effect.

### **2.3.3. Discussion**

Experiment 3 was conducted to evaluate the ecological validity of the findings revealed by Experiment 1 and Experiment 2. The main hypothesis was whether the transfer of the testing effect can be observed in a classroom setting with actual classroom materials. The results showed that the testing effect is transferable into different contexts. Students were able to use test-induced memory gain even when their memory of the material was tested in different formats (e.g., rephrased format or transformed format).

There is an interesting pattern revealed by Experiment 3: the influence of conceptual processing on the final exam. As explained above, the questions in the final test were selected from three exams throughout the semester. Half of them were conceptual and the other half were factual questions. The analysis showed that the accuracy levels of these two types of questions in the initial tests were not different,  $p > .10$ . However, in the final test, conceptual questions were answered more accurately than factual questions. Additional analyses showed that rephrasing conceptual questions yielded greater memory performance than rephrasing the factual questions. Based on this finding, one can argue that conceptual processing leads to more durable transfer than verbatim processing (as in factual questions). This finding supports the predictions of retrieval hypothesis. According to this view, the mental effort exerted in the initial task

determines how efficiently the items will be recollected in later tests. In this regard, it can be argued that conceptual questions in the initial test force students to engage in relational processing which creates more memory cues which, in turn, strengthen the memory traces for the given item (i.e., encoding variability). At the time of later recollection, an individual can use one of the available cues to retrieve the information. More interestingly, conceptual processing seems to facilitate recollection for new questions (i.e., studied-only items). The percentage of correctly studied-only conceptual questions was higher than the percentage of correctly answered factual questions. This finding implies that conceptual processing plays an important role not only for encoding but also for retrieval. At the time of testing (i.e., retrieval), encouraging students to employ relational processing might result in superior memory performance than encouraging them to simply recite the material.

## **2.4. Experiment 4**

Experiment 1, Experiment 2, and Experiment 3 tested the transfer of the testing effect in the knowledge domain and in the temporal domain. The results of the experiments showed that the testing effect is transferrable across different levels of knowledge (i.e., factual and conceptual/applied). Moreover the testing effect is transferred in the temporal domain: when the participants were tested after 1 week, as in Experiment 1 and 2, and after 12 weeks, as in Experiment 3, the testing effect was still observed. In Experiment 4, the transfer of the testing effect in the modality domain and temporal domain is investigated. The main research question is whether the testing effect changes as a result of the nature of the test (e.g., recall or recognition). One of the motives of the study is solely practical. It is essential to find out whether the testing effect is format-bound. Which type of tests generates superior testing effects?

Answering this question has implications for the behavior outside of the laboratory because learners out of the laboratory can make use of the tests accordingly.

The other reason to conduct this experiment is that the transfer of the testing effect in the modality domain composes a rather problematic research area because, as reviewed above, studies have yielded contradictory results. Some studies reported that the testing effect is obtained only when the initial test is a recall test (e.g., Carpenter et al., 2006; Glover, 1989). Other studies reported that as long as the final test is a recall test, the tests enhance memory recollection (e.g., Chan & McDermott, 2007). Furthermore, some studies reported a superior testing effect with recognition tests (Kang et al., 2007). As indicated above, some of these studies have important methodological flaws, such as absence of factorial manipulations. Experiment 4 aims to overcome this problem by using a factorial design in which participants are exposed to different types of tests (e.g., recall and recognition tests) in the initial and the final test (See Figure 9).

Other than the practical importance of the transfer of the testing effect in the modality domain, there are also theoretical concerns: the leading theories make different predictions about the transfer in the modality domain. According to the transfer-appropriate processing framework, the success of the transfer reflects the resemblance between initial test and the final test. In this respect, as the similarity increases between the tests in the two sessions, a superior transfer is expected to be observed. For example, a participant who takes a recognition test as an initial test should exhibit a superior testing effect when s/he is given a recognition test as a final test, as compared to a recall test as a final test. The retrieval hypothesis posits that testing effect is a product of higher level learning processes and is a function of retrieval attempts during the initial testing. Additionally, retrieval difficulty induced by initial testing determines the magnitude of

the testing effect. In this regard, it can be hypothesized that the retrieval hypothesis predicts that the more difficult initial test would result in superior transfer. As noted in the previous section, recall, traditionally, is accepted more difficult than recognition (for a review see Gillund & Shiffrin, 1984) because recall requires more extensive recollection of the to-be-learned information (Haist et al., 1992). It has also been proposed that recall is a product of a two-stage process in which first the item is retrieved and then a familiarity decision is made about the item (also known as the generate-recognize theory of recall). On the other hand, recognition requires only a familiarity decision since the item is already presented to the subject (Anderson & Bower, 1974). For this reason, the retrieval hypothesis would predict that an initial recall test would produce a stronger testing effect, and in turn, more robust transfer.

As indicated above, Experiment 4 also examines the transfer of the testing effect in the temporal domain. The final test is given after 1 week delay as in the previous experiments. It is expected to replicate the same findings as reported above.

### **2.4.1. Method**

#### **2.4.1.1. Participants**

Forty-nine undergraduates from the Rice University Psychology Subject Pool participated in partial fulfillment of course requirements. The data from 41 participants were analyzed in the study. The data from 8 participants were excluded due to selection criterion (See materials section for the discussion) or due to failure to follow the experimental instructions.

#### **2.4.1.2. Materials**

*Study passages.* Six essays from the journal *Monitor on Psychology* were selected as study materials. Tables and figures, if present, were removed. Please note that elimination of the

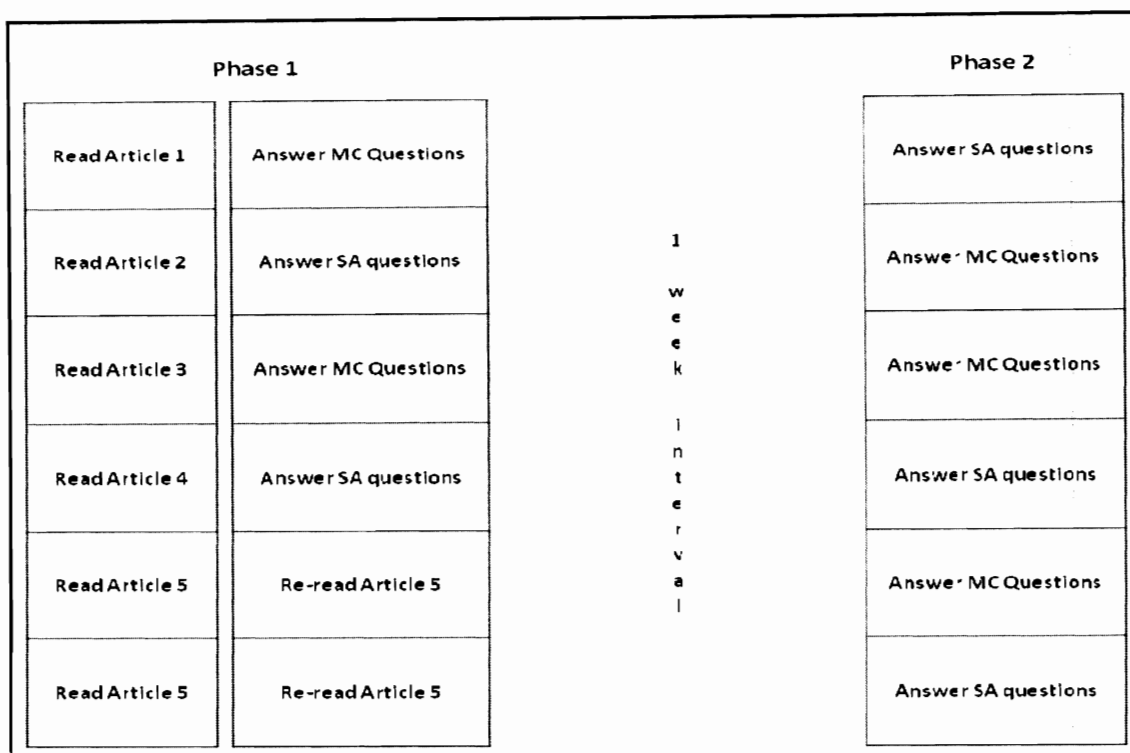
figures and tables did not affect the coherence of the essays. The average length of the essays 1,155 words ranging between 978 and 1,248 pages (See Appendix G for the essays).

*Tests.* From each essay, six facts were selected which were tested in multiple choice (MC) or short answer (SA) format. In this study, the recognition task was operationalized as a test of multiple choice questions and free recall task was operationalized as short answer questions. In the MC format, participants had to choose a response among four options, whereas in the SA format they had to generate a phrase or sentence to answer the questions (See Appendix H). The participants were also given a demographic questionnaire which aimed to assess their knowledge of the topics covered by the essays. In the demographic questionnaire (See Appendix C), the participants were asked whether they were familiar with the concepts that they read in the experiment. If the participant had any familiarity, s/he was asked to rate his/her level of knowledge on a 5-point Likert scale (1=Poor to 5=Expert). The participants with a score 3 or above were not included in the data analysis due to the possibility that their existing knowledge of the concepts could interfere with the manipulation of the experiment.

#### **2.4.1.3. Design**

A 3 (initial task: MC, SA or reread the essays) X 2 (final test format: MC or SA) within subject design was used. During the first session, participants studied the six essays. The order of the initial tasks and the order of the final task formats were counterbalanced across participants. They took initial test right after reading the essays or were given extra time to reread the essays. During the second session 1 week later, participants were tested on all of the six essays. In the final test, questions alternated between MC and SA formats. The facts tested in the final test were identical to those tested in the first session, although question format was different or same in accordance with the condition (See Figure 9 for the design of Experiment 4).





**Figure 9. The design of Experiment 4.**

#### 2.4.1.4. Procedure

Participants were tested in groups of 5 or fewer during two experimental sessions. During the first session, the participants were seated and given paper copies of the essays one at a time. At the beginning of the experimental session, participants were told that their memory of the essays would be tested. They were told not to underline or mark any part of the essays. At the onset, they were told that they would be given different types of tests after each paper, although they did not know the format of the test prior to studying the essays. They were given 10 minutes to read the each essay. After the 10 minutes elapsed, the participants were instructed to put away the paper copy of the essay. In the test condition, the participants were given 5 minutes to complete a MC test or SA test, whereas in the restudy condition, the participants were given 5 more minutes to reread the essay. Once they completed the task (i.e., tests or rereading the

essay), the participant proceeded to read the next essay. The first session lasted about 1 hour and 30 minutes.

One week later, the participants returned for the second session. They were tested on all six essays that they had studied in the first sessions. They were again tested in a MC format or SA format. After they returned the completed tests, the participants were given a demographic questionnaire which assessed their knowledge of the topics which were covered in the essays. After they filled out the demographic questionnaire, they were debriefed and dismissed. The second session lasted 20 minutes.

## **2.4.2. Results**

### **2.4.2.1. Scoring**

For MC questions, responses were counted as either correct (1 point) or incorrect (0 points). For SA questions, responses were scored as either correct (1 point), partially correct (0.5 point) or incorrect (0 points). Scoring was done by the author of the present study. As a reliability check, tests were scored by a “blind” second rater who did not know the study. The interrater reliability was .96.

### **2.4.2.2. Initial Test Performance**

Although the focus is on the final test performance, initial test performance was also examined. For each participant, the percentage of items correctly answered on the initial MC ( $M = 62.19$ ,  $SD = 17.89$ ) test and SA ( $M = 63.71$ ,  $SD = 18.12$ ) test was computed. There was no significant difference between the performance on the MC and SA tests,  $p > .05$ .

### **2.4.2.3. Final Test Performance**

The percentage of questions answered correctly by each participant from each initial test condition was computed separately for the two final test formats (MC and SA), and the resulting

descriptive statistics can be seen in Table 1. Due to scaling differences between the two formats of the final test, the final MC and SA memory performances were analyzed separately using one-way repeated measures ANOVAs with initial task as a within-subjects factor.

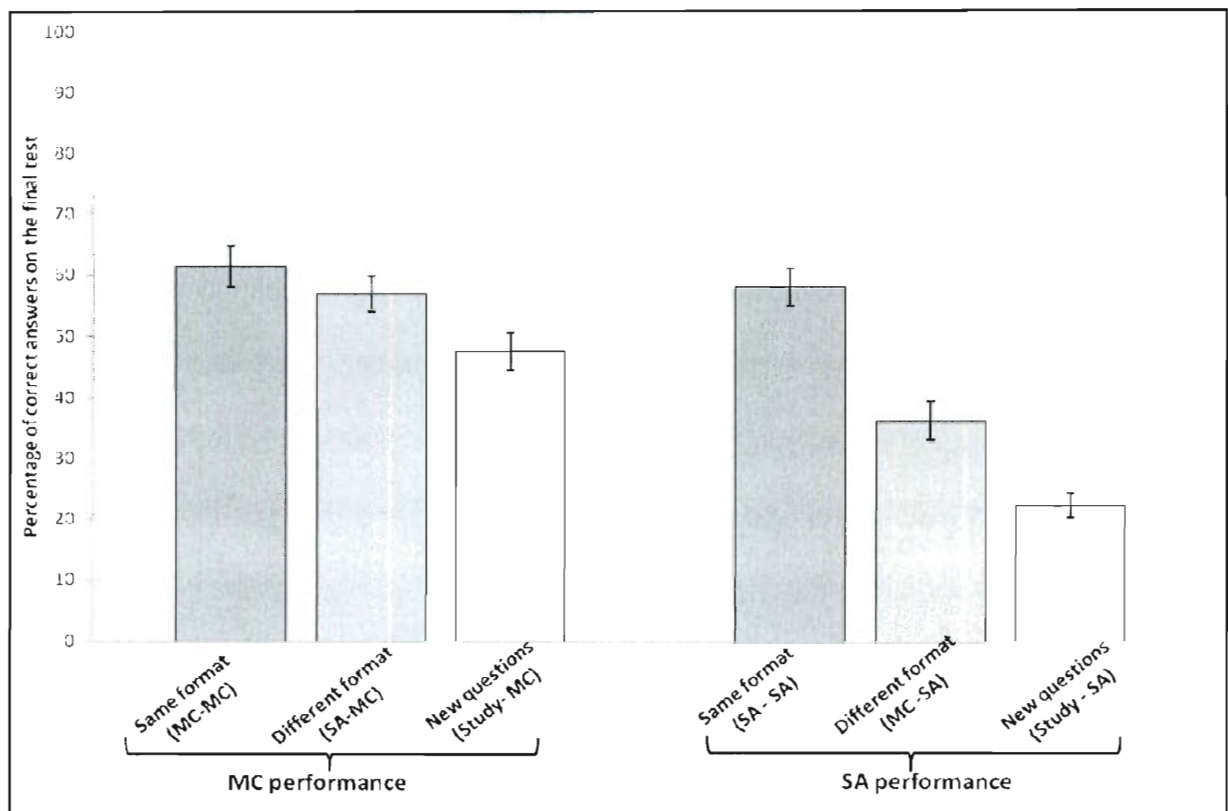
| Initial Task | Final Test Format |       |       |       |
|--------------|-------------------|-------|-------|-------|
|              | MC                |       | SA    |       |
|              | Mean              | SD    | Mean  | SD    |
| MC           | 61.38             | 21.22 | 36.18 | 20.29 |
| SA           | 56.91             | 18.19 | 57.92 | 19.02 |
| Study-only   | 47.56             | 19.2  | 22.15 | 13.28 |

**Table 1. Percentage of questions answered correctly on the final test as a function of initial task type (n = 41)**

*Multiple-choice.* The type of initial task affected final MC performance,  $F(2, 80) = 7.71$ ,  $p < .001$ ,  $MSE = 310.22$ ,  $\eta^2 = .10$ . Post hoc comparisons using paired sample t-tests showed that the initial MC test condition resulted in greater final performance than the study-only condition,  $t(40) = 3.63$ ,  $p = .001$ ,  $d = .56$ . Similarly, initial SA test condition had greater final performance than the study-only condition,  $t(40) = 3.37$ ,  $p = .01$ ,  $d = .37$ . Memory performance on the MC final test did not change as a function of the type of initial test (i.e., MC or SA),  $p < .05$ . The statistical analysis revealed a robust testing effect. Besides, the testing effect obtained in the final MC test did not differ as a function of test format in the initial test (See Figure 10).

*Short answer.* The type of initial task affected final SA performance,  $F(2, 80) = 71.13$ ,  $MSE = 187.15$ ,  $\eta^2 = .42$ . Post hoc comparisons using paired samples t-tests revealed that the

initial SA test condition had greater final performance than both the initial MC test condition and the study-only condition,  $t(40) = 6.69, p < .001, d = 1.04$  and  $t(40) = 13.09, p < .001, d = 2.04$ , respectively. The initial MC condition resulted in higher final performance than the study-only condition,  $t(40) = 3.85, p < .001, d = .60$ . The statistical analyses revealed a robust testing effect, and the testing effect gained in the final test differed as a function of the format of the initial test (See Figure 10).

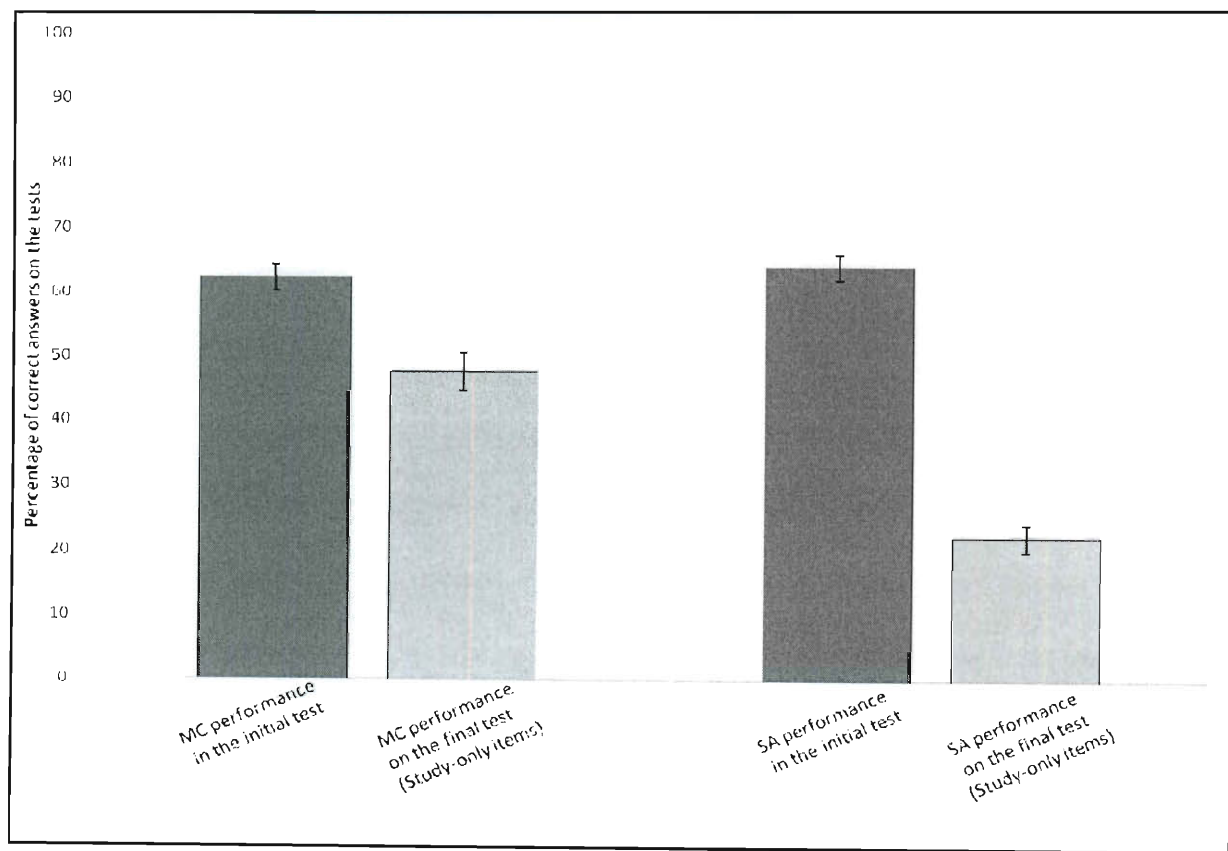


**Figure 10. The performance on the final test as a function of the match between initial task and the final task.**

#### 2.4.2.4. Initial and Final Performance

Although it is not the focus of the experiment, the forgetting rate after the 1 week delay as a function of the question format was also examined. The purpose of the analysis was to find out whether the forgetting rate changed as a result of the question format. Paired samples t-test

analyses showed that performance on the MC test declined significantly after the 1 week delay,  $t(40) = 3.77, p < .001, d = .58$ . A similar pattern was obtained in the SA test: participants' performance on the SA test after 1 week was significantly lower than performance on the immediate SA test,  $t(40) = 17.37, p < .001, d = 2.71$  ( See Figure 11). Paired sample comparisons showed that the forgetting rate was higher for SA questions ( $M = 41.56, SD = 15.32$ ) than MC questions ( $M = 14.63, SD = 24.84$ ),  $t(40) = 5.62, p < .001, d = .88$ . This finding is consistent with the argument that recall is more difficult than recognition.



**Figure 11. The forgetting rate throughout 1 week as a function of question format.**

### 2.4.3. Discussion

Experiment 4 was conducted to assess the transfer of the testing effect in the modality domain and the temporal domain. The experiment yielded the same result concerning the

transferability of the testing effect in the temporal domain: the testing effect was obtained after 1 week. In addition to the robust testing effect, the results of the statistical analyses showed that the performance on the final MC was not affected by the type of initial test whereas performance on the final SA test was susceptible to format differences in the initial test. The memory gain obtained in the SA format and the MC format in the initial test was transferable to the MC format in the final test. Transfer success did not change as a function of the format of the initial test. However, the performance on the final SA test revealed a different pattern. Although initial testing still resulted in transfer, as opposed to the study-only condition, the magnitude of the transfer rate differed as a result of initial tests. The superior transfer on the final SA test was detected when the initial test was also in SA format. On the other hand, when the initial test was in MC format, the performance on the final SA test was impaired. The results of the experiment indicate that the transfer of the testing effect in SA format is vulnerable to format differences whereas the transfer of the testing effect in MC format is more resilient to format differences.

As noted above, theories make different prediction about the transfer of the testing effect across different formats. According to the transfer-appropriate processing framework, the transfer reflects the similarity between initial test and the final test: the more resemblance between the tasks in the two sessions, the superior the transfer. The findings of the experiment provided partial support for this framework. First of all, initial testing resulted in greater final performance as compared to the study-only condition, which indicates that two tasks, initial test and the final test, are sufficiently similar to produce the transfer. The second piece of evidence supporting the prediction of this framework concerns the performance on the short answer questions. Participants' performance was superior when the initial and the final task were recall than when the initial task was recognition task and the final task was recall task. This pattern was

not obtained with recognition task, i.e., initial recognition test did not generate superior transfer when the initial task was a recognition test, which refutes the predictions of the transfer-appropriate processing. As explained above, the retrieval hypothesis predicted that an initial recall test would produce a stronger testing effect, and in turn, a robust transfer. Experiment 4 revealed that the initial recall task resulted in greater transfer only when the final task was a recall task. When the final task was a recognition task, the type of the initial test did not make a difference. The failure to find a facilitative effect of a recall task on the final recognition task deserves contemplation which is discussed in the general discussion section below.

### **3. General Discussion**

A vast array of studies have documented that retention of studied material can be enhanced by testing. Superior memory resulting from retrieval of information is known as the testing effect. The testing effect has been demonstrated with different materials (e.g., word lists, prose), with different designs (e.g., within-subjects, between-subjects), with different age groups and with different settings (e.g., laboratory, classroom) (see Roediger & Karpicke for a review).

This thesis offers a novel approach to the assessment of the testing effect: the transfer of the testing effect. Although the testing effect is a well-established phenomenon, one question that remains to be solved concerns how efficiently learners can transfer the memory gain induced by testing into different contexts. Is it transferable to diverse settings? Is the testing effect a mere result of rote-learning, which should prevent the transfer of the effect, or does it encourage a deeper level of mental processes which should facilitate the transfer of the effect? This issue is important for both practical and theoretical reasons as it is discussed below.

#### **3.1. Transfer of the Testing Effect**

Transfer of the testing effect can be evaluated in different levels and domains of transfer as described above. In the present study, the transfer of the effect in three domains was investigated: *temporal domain*, which is related to the endurance or retention of learning, *knowledge domain* where the individual transfers what he learns from one content area to another, and *modality domain*, which concerns the procedures or format of learning and testing.

The retention interval has a remarkable effect on the testing effect: as reviewed above, studies reported robust testing effect in the delayed testing conditions (e.g., Carrier and Pashler, 1992). Moreover, Roediger and Karpicke (2006) found out that the testing effect was reversed in the immediate testing condition: In the immediate final memory test, studied-only items were recalled/recognized better than tested items. In the present study, the transfer of the testing effect in the temporal domain was examined by using a 1 week interval in Experiment 1, Experiment 2 and Experiment 4, and up to 3 months in Experiment 3. The results of the experiments reported here provided evidence that the testing effect endures over 1 week: participants can utilize the memory gain produced by initial tests after a 1 week interval. Furthermore, the results of Experiment 3 showed that the effect can last as much as 3 months. Based on this finding, it can be derived that tests promise long lasting enhancing effects on memory. These and similar findings point out that the testing effect is transferable across different temporal domains.

The transfer of testing effects in the knowledge domain was examined in two dimensions: near and far transfer. Near transfer was operationalized as follows: the questions in the initial tests were rephrased in the final tests. Experiment 1 showed the testing effect was evident with rephrased questions: Participants showed a test-induced memory gain when they were tested with rephrased questions, which indicates that participants' memory gain resulting from the initial test is transferable. However, the apparent superiority of the testing effect resulting from



identical questions in the final test indicates that rote-learning plays a role in the testing effect as well. To evaluate far transfer in the knowledge domain, the questions in the initial test were transformed into different levels of knowledge. To achieve this goal, as described above in the introduction section, Bloom's (1956) taxonomy was used as a guide. Bloom defined six levels of the cognitive domain: knowledge, comprehension, application, analysis, synthesis, and evaluation. Knowledge level represents the lowest level of learning where a learner memorizes, defines or duplicate the learned material. In each ascending level, the learning processes require more complex and abstract mental levels. In Experiment 2 and Experiment 3, knowledge level questions, which are labeled as *factual questions* in the present study, were transformed in to application level questions, which are labeled as *conceptual questions*. Application level questions aim to force the students to use the acquired knowledge in different ways to solve new questions. The conceptual questions are constructed solely for this purpose: to compel the participants to use the memory gain in different contexts which requires greater comprehension of the material rather than simply memorizing it. The results of the Experiment 2 and Experiment 3 showed that on the final test, participants' performance on the factual and conceptual questions were compatible which indicated that participants could transfer the knowledge of specific facts into the application level and vice versa. However, when the initial test scores were taken into account, a different picture emerged. When the participants are exposed to factual questions in the initial test, the forgetting rate was higher, especially when the final test was composed of conceptual questions. In other words, the memory gain induced by initial factual questions is more vulnerable to decay, hence more resistant to transfer. On the other hand, initial conceptual questions generated a testing effect which was more robust and resistant to forgetting. Although in the initial test, conceptual questions resulted in lower memory performance than factual

questions, in the following tests the participants utilized the memory gain to the greater extent than the memory gain induced by factual questions. This pattern precisely characterizes the difference between knowledge level and higher levels of learning in Bloom's taxonomy. As adequately put forth by Mazur (2008), in a video clip where he explains the use of *ConceptTest*, "you can forget facts, but you can't forget understanding." Practical importance of conceptual learning was emphasized by Manzur (1997) who coined the term *ConceptTest*. Manzur is physics professor at Harvard University. As he realized that students were passing the exams without understanding the concepts, he developed *ConceptTests* which are multiple-choice questions asked during the lecture and cannot be solved by memorization of the equations. For example, one question which was developed by McConnell et al. (2006) solely for this purpose is as follows:

Mars has a more eccentric orbit of the Sun than Earth. Mars is 20% closer to the Sun during its winter than it is during its summer. How would temperature distributions on Earth be affected if we had a similarly eccentric orbit that brought us much closer to the Sun during winter in the Northern Hemisphere?

- a. Maximum temperatures would be higher during the winter season in the Northern Hemisphere.
- b. Maximum temperatures would be higher during the winter season in the Southern Hemisphere.
- c. Maximum temperatures would be higher during the winter seasons in both the Northern and Southern Hemispheres.

Later on, *ConceptTests* were adapted to higher level of domains in Bloom's taxonomy, such as comprehension and application levels. The facilitative effect of *ConceptTests* on learning

was reported in an experimental study by McConnell et al. (2006) who compared two classes which were taught during different semesters by the same instructor. The first class was taught without using any ConcepTests whereas the second class utilized ConcepTests. Note that both of the classes covered the same materials. The students exhibited greater gains in the ConcepTests incorporated class. Combining these studies with the results of the present study reveals that encouraging students to engage in higher order learning promotes comprehension of the material results in greater transfer. The same pattern is applicable to the transfer of the testing effect. In the initial tests, it is of essential important to induce the conceptual processing of the materials which promotes testing effects to a greater extent than mere memorization of the facts and terms. Many believe that promoting longer lasting and transferable learning should be the primary goal of all educational systems (e.g., Barnett and Ceci, 2002). It seems that the testing effect can be employed to achieve this goal.

The transfer of testing effects in the modality domain was investigated by manipulating different formats of tests: recall and recognition. Experiment 4 reported that the testing effect is transferable across different formats albeit to a different degree. Recognition performance was not affected by the format of the initial test whereas recall performance was the greatest when the initial test was also a recall task. This pattern can be interpreted in two ways. One way of explanation concerns that vulnerability of recall tasks to format differences: Recall tasks are affected by different formatting: higher testing effects in the final test are attained when the initial test is also a recall task. The second way of interpreting this pattern is related to the power of recognition tasks. It can be speculated that the reason why the transfer was weaker in the final recall task when the initial task test a recognition task might lie in the fact that recognition tasks are not strong enough to promote transfer effect. Thus, taking an initial recognition test does not

do much to help transfer of learning onto a final recall test, as seen in Experiment 4. Recognition tasks are known to be easier than short recall tasks (Gay, 1980) and more difficult tasks enhance learning than easier tasks (Sax & Reade, 1964) which promotes transfer of knowledge (Schmidt & Bjork, 1992). Thus, perhaps the most sensible conclusion to be made from this data and others is that multiple choice tests, as an example of recognition task, do not promote transfer as much as short answer tests. This argument has important theoretical implications as are discussed below.

To sum up, the present study demonstrated that the testing effect relies more heavily on deeper levels of mental processes which have been substantiated by the transferability of the testing effect in three domains as described above.

### **3.2. Theoretical Implications**

On the theoretical front, the transfer of the testing effect provides insight into the mechanisms underlying the testing effect. As reviewed above, each of the theories makes different predictions about the transferability of the testing effect. To summarize very briefly; the retrieval hypothesis predicts significant transfer effects due to the fact that this framework views the testing effect as a result of deeper cognitive processes which increases the probability of the transferability of the testing effect in all of the domains of interest. The transfer-appropriate processing framework makes very specific predictions about the transfer: as long as the initial and the final processes are compatible, the transfer is expected to occur. In this section, the findings of the study will be discussed as they relate to the theoretical orientations.

The transfer of the testing effect in the temporal domain is predicted by both the retrieval hypothesis and transfer-appropriate hypothesis, albeit for different reasons. The mental processes underlying the testing effect as proposed by the retrieval hypothesis should support long term

retention of the testing effect. The frameworks assert that the act of retrieving information increases the accessibility to the retrieved information due to proliferation of the memory cues and strengthening of the memory traces (e.g., Carpenter & DeLosh, 2006). The testing effect, for this framework, is rooted beyond rote memorization. It can be safely speculated that information relying on multiple memory cues and strengthened memory traces should be more enduring than memorized information. On the other hand, the transfer-appropriate hypothesis emphasizes the resemblance between the initial task and the final task for the occurrence of any transfer. As long as the initial and final processes are similar, the retention of information should be superior. In this regard, this framework predicts that testing effects would be carried throughout long retention interval as reported in the present study. In this regard, the results of the experiments reported here provide supporting evidence for both of the theories. In other words, the methodology employed in this study cannot differentiate between two theories since both of the theories predict the enduring transfer effect.

On the other hand, the theoretical orientations within the scope of the present study make different predictions about the transfer of the testing effect in the knowledge domain. In this study, the transfer in the knowledge domain was manipulated in different levels: near and far transfer. Near transfer was operationalized as rephrasing of the questions in the final test whereas far transfer was operationalized as the transformation of the questions in the final test (e.g., factual vs. conceptual). The retrieval hypothesis predicts that the effect should be transferred across different levels of the knowledge domain independent from the nature of the questions whereas the transfer-appropriate processing framework predicts that the resemblance between the initial and final test environments determines the success of the transfer. Experiment 1 and Experiment 3 reported a near transfer effect of the testing effect: memory gains produced by

initial testing were seen in the final tests even when the questions were rephrased. This finding was predicted by the retrieval hypothesis. However, memory performance was still best in the identical wording conditions. The evident advantage of identical wording indicates that the testing effect also reflects compatible processes at the time of encoding and retrieval which supports the predictions of the transfer-appropriate processing. As described in the method sections above, in the same wording condition, participants were exposed to the same questions in the initial and final tests. From the perspective of transfer-appropriate processing view, this condition is the perfect example of very similar processes in the encoding and retrieval conditions which, in turn, generated the greatest testing effect as compared to other conditions. As the similarity decreased between the encoding and the retrieval conditions, as in the rephrased condition, the testing effect weakened. Although, the finding concerning near transfer of the testing effect seems to support the predictions of the transfer appropriate processing view, far transfer in the knowledge domain also yielded interesting pattern. The transfer effect obtained with the transformed questions and non-transformed questions was compatible which indicates that participants could transfer memory gains induced in the initial tests to their final test performance independent of the type of the questions. This finding supports the predictions of the retrieval hypothesis: testing effects reflect deeper memory processes and are transferable to different contexts. The transfer-appropriate processing view, on the other hand, predicts that the transfer should be greatest in the compatible conditions. For example, the transfer should be superior in Factual-Factual and Conceptual-Conceptual conditions than in Factual-Conceptual and Conceptual-Factual condition. This is because, in latter two conditions, the initial and final tests draw on different processes. However, the findings reported in this study failed to confirm the prediction of this framework. Indeed, initial factual questions produced weaker testing effects

whereas the initial conceptual questions produced stronger testing effects as measured by forgetting rates after 1 week retention interval. This finding is predicted by the retrieval hypothesis. As the difficulty of the encoding process increases, the mental elaboration on the material also increases (Sax & Reade, 1964). This situation is described as “desirable difficulties” by Bjork (1994) as explained in the introduction section. In this regard, using tests as a means of encoding introduces optimal difficulty for learners for later successful retention (Roediger & Karpicke, 2006a). Moreover, conceptual questions, in this study, create even more difficult learning environment as indicated by the decreased recognition rates in the initial conceptual questions compared the factual questions as proposed by Bloom. Although difficult learning conditions (conceptual questions) dampened memory performance in the initial test, recognition rates were better on the final test after 1 week. This finding implies that factual questions facilitate performance during the acquisition phase but are ineffective for learning as measured on a transfer test. Conceptual questions displayed an opposite pattern as described above. This finding seems to support strongly the retrieval hypothesis: introducing difficulty in the acquisition phase promotes better retention. The overall results indicate that although the transfer appropriate framework is supported partially, it is not supported in crucial ways: (1) similarity between initial and final test did not affect the strength of testing effect and (2) initial factual questions produced weaker transfer whereas the initial conceptual questions produced stronger transfer. These two crucial findings support the predictions of the retrieval hypothesis.

The transfer of the testing effect in the modality domain, as discussed above, is a rather problematic area of research since several studies reported conflicting findings. In this study, it was found that the testing effect is transferable across different formats albeit, to a different degree. The findings of this study partially support the predictions of the transfer-appropriate

processing view. Superior recollection found in testing conditions (testing effect) compared to study-only conditions reflects the idea that the two tasks, initial test and the final test, are sufficiently similar to produce the transfer. Participants' superior performance in the condition where the initial and the final task were recall than in the condition where the initial task was recognition task and the final task was recall task also support the transfer-appropriate processing view as similar contexts of learning and testing condition boosted memory performance. However, this pattern was not obtained with recognition task, i.e., initial recognition test did not generate superior transfer when the initial task was a recognition test, which refutes the predictions of the transfer-appropriate processing.

Recognition performance was not affected by the format of the initial test whereas recall performance was the greatest when the initial test was also a recall task which suggests that recall tasks are more susceptible to the changes in formats than recognition tasks. At first glance, the recognition task seems to be more efficient for transfer purposes as it is not affected by changes in the format of the initial test. However, from another point of view, by looking at the Figure 10, it can be argued that the initial recall test boosts the recollection rates in the final test more than the initial recognition task: the initial recall task produced higher recognition and recall performance in the final test whereas the initial recognition task produced higher recognition performance yet failed to improve the recall performance in the final recall test. From this point of view, the prediction of the retrieval hypothesis was confirmed: a more difficult initial task produced greater testing effects. This trend is also predicted by a dual processing account. In terms of Jacoby's (1991) dual processing account, the reason for this differential effect might be due to different processes involved in recall and recognition. According to this model, short answer tests rely on deeper engagement of recollection which



leads to stronger transfer as compared to multiple choice tests which rely on familiarity, which in turn creates less positive transfer to a final short answer test.

The overall pattern, once again, provides partial support for the transfer-appropriate processing view. However, the data in the study do not support this framework in crucial ways – namely the fact that recall produces the best performance no matter what the final test is (e.g., whether it matches the initial test or not).

### **3.3. Practical Implications**

Recently, in educational settings, traditional teaching practices are being questioned. Instead of passive teaching practices, instructors are more inclined to use more active techniques to foster learning (e.g., Mayer et al., 2009). For example, in a recent article, Felder and Brent (2009) stated the essential role of active learning in the classrooms to promote learning. They defined active learning as “anything course-related that all students in a class session are called upon to do other than simply watching, listening and taking notes” (p. 2). They argue that active learning promotes better retention than sitting passively in the classroom. Their argument has been supported by several pieces of evidence coming from research and in-class practices. For example, *ConceptTests*, as explained above, are considered to be an example of active learning which improved the retention of information (McConnell et al., 2006).

The present study also proposes that tests, which have traditionally been used as an assessment means, can also be used in the classroom to promote learning. In recent years, the testing effect is much more appreciated not only in academia as a research field, but also in educational settings. Instructors are realizing that they can utilize tests as an important means of learning (e.g., Haynie, 1994; Marshak and Logan, 2010). In fact, it has been reported that frequent quizzing has been found to lead to higher performance on later exams (e.g., Truell,

Cassady & Jameson, 2008). Bangert-Drowns et al. (1991) conducted a meta-analysis study which pooled 35 studies. The most important feature of this meta-analysis was that it included studies which took place in real classrooms; laboratory studies and research using paid volunteers were not included in the study. In this regard, the meta-analysis aimed to evaluate quizzing in the actual classroom setting. Twenty nine of the 35 studies reported positive effects from frequent testing with an average effect size of .23. The authors also reported the students' attitude toward classes after they were exposed to tests. Students, in the frequent testing conditions, rated their classes more favorably than the students who were less frequently tested. Tests are not only used to improve retention but also are used to increase class participation and preparation (Marcell, 2008). Marcell (2008) reported that taking a quiz on daily readings increased the number of questions and comments in the class and encouraged the students to read the assigned materials. As it is seen, tests can be used to encourage students to develop efficient studying habits.

The results of the present study also showed the strength of the testing effect by examining the transferability of the testing effect in different domains. In addition to the demonstration of the fact that the tests can be a powerful tool for supporting long-lasting and flexible learning, one of the most crucial implications of the findings of the present study is the importance of introducing optimal difficulty into tests. Conceptual processing, as manipulated by tests including conceptual questions, resulted in superior memory which was resistant to decay after 1 week retention interval. Although factual processing gave rise to superior memory in immediate testing, it was subject to greater decay after 1 week retention interval. Moreover, the facilitative effect of conceptual processing expanded to factual questions after 1 week as shown in Figure 6: forgetting rates were lower for factual questions in the final test when the

participants were given a conceptual test as an initial test. This effect was not replicated with an initial factual test. Based on these results, it can be derived that conceptual processing has a broader general facilitative effect than factual processing. In this respect, the present study extended the findings of a study by Hamaker (1986) who examined the effect of adjunct questions on prose learning. Adjunct questions are defined as “questions added to an instructional text to influence what is learned from the text” (Hamaker, 1986, p.212). The difference between adjunct questions and tests as used in the testing effect studies is that adjunct questions are given before to-be-learned text, and the participants are asked to look for the answers of the questions as they read the text while in testing effect studies, the test is given after reading the materials. Hamaker (1986) manipulated two types of questions: factual and higher order. Factual questions ask the students to repeat or recognize information as it was provided in instruction whereas higher order questions require students to manipulate the learned information to create an answer. As can be seen, the operationalization of the types of question is very similar to operationalization of the questions in the present study. Hamaker (1986) found that although factual adjunct questions facilitated the learning of the material, the effect was not as great as higher order questions.

The present study also showed that the facilitative effect of conceptual processing is not limited to the testing effect. Figure 5 and Figure 8 reveal that memory for newly tested information can be enhanced by encouraging students to engage in conceptual processes (as measured through performance on the new questions on the final tests).

The importance of optimal difficulty in tests has been also shown in Experiment 4. Short answer (SA) questions facilitated the performance with multiple-choice (MC) questions after 1 week. However, MC questions did not exert such an effect on SA questions after 1 week. Since

SA questions are accepted to be more difficult than MC questions, it can be speculated that the difficulty induced in the initial test is responsible for such transfer.

### **3.4. Future Directions**

The testing effect is a robust phenomenon as demonstrated in the present study and several other studies. Compared to one decade ago, we know much more about testing and why it can work as a learning strategy, not just a way to factor grades for a class. The facilitative effect of tests, promisingly, is no longer limited to laboratory settings. Instructors have started to systematically utilize tests as teaching and learning tool. However, there is still much more to explore in testing effect phenomena.

One of the issues yet to be investigated concerns the optimum amount of testing as raised by Banger-Drowns et al. (1991). As reviewed in the present study, cognitive researchers argue that tests promote the rehearsal of newly learned information which prevents the decay of memory. On the other hand, researchers in the educational literature have been skeptical about the effects of testing. For example, Proger and Mann (1973) published an extensive literature review of research on the test as a learning device. Based on the review, they argued that the reports are very contradictory to propose any facilitative effect of frequent testing. However, recent developments in experimental research and well-controlled studies have repeatedly shown that testing, as reviewed here, promotes learning. However, the biggest concern of educational research seems to be the frequency of testing. Some educators argue that focusing on exams might direct students to perform better on the test rather than to learn the materials (e.g., “teaching the test”, Banger-Drowns et al., 1991). Also some educators state that too frequent testing takes time away from instruction and discussion (Roediger, McDaniel & McDermott, 2006). To utilize tests as a learning means in the classroom, the focus should be put on this

issue. To my knowledge, there is only one study which investigated the gains of repeated testing in the classroom and one in the laboratory study. As summarized above, Banger-Drowns et al. (1991) conducted a meta-analysis study which pooled 35 in-class studies. In addition to the facilitative effect of tests in the classroom, they also investigated the effect of increasing test frequency. They found out that memory gains are incrementally smaller with each test. Their regression analyses revealed that when the teachers included one exam to their course plan, the performance increased .34 standard deviations. Adding a second test increased final exam scores by only .08 standard deviations. The similar finding was reported by Roediger and Karpicke (2006b) who, in Experiment 2, assessed the testing effect in three conditions (S = Study, T = Test): repeated study condition (SSSS), single test condition (SSST) and repeated test condition (STTT). Results revealed that SSST and STTT condition resulted in the highest correct recall. There was no significant difference between these two conditions. Based on this finding, it can be speculated that there is an optimum amount of testing to enhance memory as reported by Banger-Drowns et al. (1991). This finding raises thought-provoking questions about the frequency of tests in the classrooms and deserves experimental research.

Another topic that deserves further attention is the phenomenological experience of learners at the time of testing. Phenomenology is defined as “the study of structures of consciousness as experienced from the first-person point of view. The central structure of an experience is its intentionality, its being directed toward something, as it is an experience of or about some object” by Stanford Encyclopedia of Philosophy. Phenomenological experience as it applies to the testing effect concerns what the learners experience when they are given tests for encoding purposes. One study, indirectly, investigated this issue. Pyc and Rawson (2010) asked students to learn Swahili words by matching them with their English counterparts. Half of them

were given a practice test after studying the word pairs and the other half restudied the words. The result displayed a robust testing effect: the practice test group remembered more words than the restudy group. The importance of this study comes from the fact that the researchers asked the participants to tell them their keywords to learn the word pairs. They found out that participants in the testing condition remembered more of the keywords. Moreover, the participants were more likely to change their keywords as they progressed in the learning phase. This study paves the way for examining individual experience of testing effect. This approach not only informs us about the dynamics of the testing effects but also allows us to tailor tests in accordance with the individual differences.

The third issue which deserves further research is the role of feedback in the testing effect. The role of feedback in learning and memory is an extensive topic which merits another dissertation of research by itself. As it relates to the testing effect, the role of feedback begs further exploration. In the present study, feedback was not given the students for two reasons. First, feedback was not found to affect testing effects as a result of multiple choice questions (Kang et al., 2007). In the present study, except Experiment 4, multiple choice questions were used. The second reason concerns the strength of the testing effect. As has been reported repeatedly, the testing effect is so strong that even without feedback it outperforms the study-only condition. To prevent an unfair advantage to the participants in the testing condition, the feedback was eliminated from the study. However, it does not mean that the role of feedback should not be investigated. On the contrary, even in the absence of feedback, the testing effect manifests itself very strongly. However, as noted above, multiple choice questions are not affected by feedback. On the other hand, two studies reported that feedback has a differential effect on the testing effect depending on the format. Pashler, Cepeda, Wixted and Rohrer (2005)

in a cued recall task and Kang et al. (2007) in a recall task showed that feedback enhances the testing effect. This issue begs for additional research.

The present study utilized a novel way for assessing the strength of the testing effect: Can memory gain as a result of taking test be transferred to different domains? In other words, I aimed to learn how just powerful the testing effect is. The results revealed that the testing effect is mostly independent of the nature of the to-be-learned materials. A learner can utilize test-induced memory gains even when s/he is asked in different formats. Moreover, during the learning process, increasing the difficulty of tests promotes enduring retention which employs the well-known idiom “Easy come, easy go” into the memory and learning domain.

## References

- [1] Abbott, E. E. (1909). On the analysis of the factors of recall in the learning process. Psychological Monographs, 11, 159–177.
- [2] Auble, P. M., & Franks, J. J. (1978). The effects of effort toward comprehension on recall. Memory & Cognition, 6, 20–25.
- [3] Baddeley, A. D. & Dale, H. C. A (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. Journal of Verbal Learning and Verbal behavior, 5, 417-420.
- [4] Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. L. C. (1991). Effects of frequent classroom testing. Journal of Educational Research, 85, 89-99.
- [5] Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological-Bulletin, 128, 612-637.
- [6] Begg, I. , Snider, A. , Foly, F. & Goddard, R. (1989). The generation effect is no artifact: Generating makes words distinctive. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 977-989.
- [7] Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). Cambridge, MA: MIT Press.
- [8] Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. Essex, England: Harlow.
- [9] Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 361-392). Hillsdale, NJ: Lawrence Erlbaum Associates



- [10] Bransford, J., Brown, A., & Cocking, R. (Eds.) (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.
- [11] Brewer, M. (2000). Research Design and Issues of Validity. In Reis, H. and Judd, C. (eds) Handbook of Research Methods in Social and Personality Psychology. Cambridge:Cambridge University Press.
- [12] Butler, A. C., & Roediger, H. L., III (2007). Testing improves long-term retention in a simulated classroom setting. European Journal of Cognitive Psychology, 19, 514-527.
- [13] Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, 35, 1563-1569.
- [14] Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. Memory & Cognition, 34, 268-276.
- [15] Carpenter, S., Pashler, H., & Vul, E. (2007). What Types of Learning are Enhanced by a Cued Recall Test? Psychonomic Bulletin & Review, 13, 826 -830.
- [16] Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. Memory and Cognition, 20, 632-642.
- [17] Chan, J. C. K. & McDermott, K. B. (2007). The effects of frontal lobe functioning and age on veridical and false recall. Psychonomic Bulletin & Review, 14, 606-611.
- [18] Chan, J.C.K., McDermott, K.B, & Roediger, H.L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. Journal of Experimental Psychology: General, 135, 553-571.

- [19] Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic memory. Psychological Review, 82, 407-428.
- [20] Duchastel, P. C. (1979). Retention of prose materials: The effects of testing. Journal of Educational Research, 72, 299-300.
- [21] Duchastel, P. C. (1981). Retention of prose following testing with different types of test. Contemporary Educational Psychology, 6, 217-226.
- [22] Felder, R. M., & Brent, R. (2009). Active learning: An introduction. ASQ Higher Education Brief, 2, 1- 5.
- [23] Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. Memory & Cognition, 1, 213-216.
- [24] Gates, A. I. (1917). Recitation as a factor in memorizing. Archives of Psychology, 6(40).
- [25] Gay, L. R. (1980). The comparative effects of multiple-choice versus short-answer tests on retention. Journal of Educational Measurement, 17, 45-50.
- [26] Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recall and recognition. Psychological Review, 91, 1-67.
- [27] Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. Journal of Educational Psychology, 81, 392-399.
- [28] Godden, D. R. & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. British Journal of Psychology, 66, 325-331.
- [29] Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 18, 691-702.

- [30] Halamish, V. & Bjork, R. A. (2010, November). Using tests to overcome retroactive interference. Poster session presented at the annual meeting of the Psychonomic Society, St. Louis, MI.
- [31] Hamaker, C. (1986). The effects of adjunct questions on prose learning. Review of Educational Research, 56, 212–242
- [32] Haynie, W. J. (1994). Effects of multiple-choice and short answer tests on delayed retention learning. Journal of Technology Education, 6, 32-44.
- [33] Hirshman, E. L., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 14, 484-494.
- [34] Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. European Journal of Cognitive Psychology, 19, 528-558.
- [35] Karpicke, J. D. & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. Journal of Memory and Language, 62, 227-239.
- [36] Larkin, J. H. (1989) What kind of knowledge transfers? In Resnick, L. B. (Ed.), Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser (p. 283-306). Hillsdale, NJ: Lawrence Earlbaum Associates.
- [37] Lee, T. D. (1988). Transfer-appropriate processing: A framework for conceptualizing practice effects in motor learning. In O.G. Meijer & K. Roth (Eds.), Complex motor behavior: The motor-action controversy ( p. 201-215). Amsterdam: Elsevier Science.
- [38] Little, J. L. & Bjork, E. L. (2010, November). Pre-testing with multiple-choice questions facilitates learning. Poster session presented at the annual meeting of the Psychonomic Society, St. Louis, MI.

- [39] Logan J. M., & Balota, D. A. (2008). Expanded versus equal interval spaced retrieval practice: Exploring different schedules of spacing and retention interval in younger and older adults. Aging, Neuropsychology, and Cognition, 15, 257-280.
- [40] Marcell, m. (2008). Effectiveness of regular online quizzing in increasing class participation and preparation. International Journal for the Scholarship of Teaching and Learning, 2, 1-9.
- [41] Marshak, D., & Logan, J. (April, 2010). The use of repeated practice in teaching human anatomy. Poster session presented at the annual Teagle Collegium on Student Learning, NewYork City, NY.
- [42] Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. Contemporary Educational Psychology, 34, 51-57.
- [43] Mazur, E. (Harvard's Derek Bok Center). (2008). From questions to concepts: Interactive teaching in physics. Available from <http://bokcenter.harvard.edu>
- [44] Mazur, E. (1997). Peer instruction: A user's manual. Upper Saddle River, N.J.: Prentice Hall.
- [45] McConnell, D. A., Steer, D. N., Owens, K. D., Knott, J. R., Van Horn, S., Borowski, W., ... Heaney, P. J. ( 2006). Using conceptests to assess and improve student conceptual understanding in introductory geoscience courses. Journal of Geoscience Education, 54, 61-68.
- [46] McDaniel, M. A., Anderson J. L., Derbish M. H., & Morisette, N. (2007). Testing the testing effect in the classroom. European Journal of Cognitive Psychology, 19, 494-513.

- [47] McDaniel M. A., & Masson M. E. J. (1985). Altering memory representations through retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition , 11, 371–385.
- [48] McFarland, C. E., Jr., Rhodes, D. D., & Frey, T. J. (1979). Semantic-feature variability and the spacing effect. Journal of Verbal Learning and Verbal Behavior, 18, 163- 172.
- [49] Modigliani, V., & Hedges, D. G. (1987). Distributed rehearsals and the primacy effect in single-trial free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 426-436.
- [50] Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer-appropriate processing. Journal of Verbal Learning and Verbal Behavior, 16, 519–533.
- [51] Nungester, R. J. & Duchastel, P. C. (1982). Testing versus review: Effects on retention. Journal of Educational Psychology, 74, 18–22.
- [52] Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? Journal of Experimental Psychology: Learning, Memory, and Cognition, 31, 3-8.
- [53] Pastorino, E. E. & Doyle-Portillo, S. M. (2009). What is psychology? (2<sup>nd</sup> Edition). Baltimore: Thomson Wadsworth.
- [54] Proger, B. B., & Mann, L. (1973). A historical review of theoretical and experimental literature on the teaching values of informal (non-standardized), teacher-made achievement tests: 1913-1968. Blue Bell, PA: Montgomery County Intermediate Unit23. (ERIC Document Reproduction Service No. ED 084 292)
- [55] Pyc, M. A. & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? Journal of Memory and Language, 60, 437-447.

- [56] Rajaram, S., Srinivas, K., & Roediger, H. L. (1998). A transfer-appropriate account of context effects in word fragment completion. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 993-1004.
- [57] Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science, 1, 181-210.
- [58] Roediger, H. L. & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17, 249-255.
- [59] Roediger, H. L., McDaniel, M., & McDermott, K. (2006). Test-enhanced learning. APS Observer, 19, 28.
- [60] Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 233-239
- [61] Runquist, W. N. (1983). Some effects of remembering on forgetting. Memory & Cognition, 11, 641-650.
- [62] Sax, G., & Reade, M. (1964). Achievement as a function of test difficulty level. American Educational Research Journal, 1, 22-25.
- [63] Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science, 3, 207-217
- [64] Simons, P. R. J. (1999). Transfer of learning: paradoxes for learners. International Journal of Educational Research, 31, 577-589.
- [65] Thomas, A.K., & McDaniel, M.A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. Memory & Cognition, 35, 668-678.

- [66] Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. Journal of Experimental Psychology: Human Learning and Memory, 4, 210–221
- [67] Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. Experimental Psychology, 56, 252-257.
- [68] Truell, A. D., Cassady, J. C., & Jameson, M. (2008). The impact of practice computer-based assessments on business student test performance. Journal of Business and Training Education, 17, 1 - 7.
- [69] Tulving, E., & Thomson, D. M., (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.
- [70] Watkins, M. J. & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. Journal of Verbal Learning and Verbal Behavior, 18, 687-704.
- [71] Wenger, S. K., Thompson, C. P., & Bartling, C. A. (1980). Recall facilitates subsequent recognition. Journal of Experimental Psychology: Human Learning and Memory, 6, 545-559.
- [72] Wheeler, M. A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. Memory, 11, 571–580.
- [73] Wixted, J. T. (1990). Analyzing the empirical course of forgetting. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 927-935.
- [74] Woodworth, R. S., & Scholosberg, H. (1954). Experimental psychology. New York: Holt
- [75] Yuille, J. C. (1973). Changes in verb meaning and the sentence context effect in children's PA learning. Journal of Experimental Child Psychology, 15, 247-253.

## Appendix A

### The Articles Used in Experiment 1

#### ATTRIBUTION THEORY

Attribution theory (Weiner, 1980, 1992) is probably the most influential contemporary theory with implications for academic motivation. It incorporates behavior modification in the sense that it emphasizes the idea that learners are strongly motivated by the pleasant outcome of being able to feel good about themselves. It incorporates cognitive theory and self-efficacy theory in the sense that it emphasizes that learners' current self-perceptions will strongly influence the ways in which they will interpret the success or failure of their current efforts and hence their future tendency to perform these same behaviors.

According to attribution theory, the explanations that people tend to make to explain success or failure can be analyzed in terms of three sets of characteristics: First, the cause of the success or failure may be internal or external. That is, we may succeed or fail because of factors that we believe have their origin within us or because of factors that originate in our environment. Second, the cause of the success or failure may be either stable or unstable. If the we believe cause is stable, then the outcome is likely to be the same if we perform the same behavior on another occasion. If it is unstable, the outcome is likely to be different on another occasion. Third, the cause of the success or failure may be either controllable or uncontrollable. A controllable factor is one which we believe we ourselves can alter if we wish to do so. An uncontrollable factor is one that we do not believe we can easily alter.

Note that this factor is distinct from the previous two categories. An internal factor can be controllable (we can control our effort by trying harder) or uncontrollable (most people cannot easily change their basic intellectual ability or change from being an introvert to being an extrovert). Likewise, an external factor can be controllable (a person failing a difficult course could succeed by taking an easier course) or uncontrollable (if calculus is difficult because it is abstract, it will still be abstract no matter what we do).

An important assumption of attribution theory is that people will interpret their environment in such a way as to maintain a positive self-image. That is, they will attribute their successes or failures to factors that will enable them to feel as good as possible about themselves. In general, this means that when learners succeed at an academic task, they are likely to want to



attribute this success to their own efforts or abilities; but when they fail, they will want to attribute their failure to factors over which they have no control, such as bad teaching or bad luck.

The basic principle of attribution theory as it applies to motivation is that a person's own perceptions or attributions for success or failure determine the amount of effort the person will expend on that activity in the future.

There are four factors related to attribution theory that influence motivation in education: ability, task difficulty, effort, and luck. In terms of the characteristics discussed previously, these four factors can be analyzed in the following way: Ability is a relatively internal and stable factor over which the learner does not exercise much direct control. Task difficulty is an external and stable factor that is largely beyond the learner's control. Effort is an internal and unstable factor over which the learner can exercise a great deal of control. Luck is an external and unstable factor over which the learner exercises very little control.

Note that it is the learner's perception that determines how attributions will influence future effort. A learner may believe that he is a "lucky person" - and for him luck would be an internal and stable characteristic over which he exercises little control. In other words, for this person "luck" is really what the preceding list calls an "ability" or personality characteristic. Likewise, a person may believe that she expended a great deal of effort, when in fact she did not, or that an objectively easy task was difficult. The basic principle of attribution theory as it applies to motivation is that a person's own perceptions or attributions for success or failure determine the amount of effort the person will expend on that activity in the future.

## GESTALT PSYCHOLOGY

Organizing raw sensory stimuli into meaningful experiences involves cognition, a set of mental activities that includes thinking, knowing, and remembering. Knowledge and experience are extremely important to perception, because they help us make sense of the input to our sensory systems.

How people perceive a well-organized pattern or whole, instead of many separate parts, is a topic of interest in Gestalt psychology. According to Gestalt psychologists, the whole is different than the sum of its parts. *Gestalt* is a German word meaning *configuration* or *pattern*.

A major goal of Gestalt theory in the 20th century was to specify the brain processes that might account for the organization of perception. Gestalt theorists, chief among them the

German-U.S. psychologist and philosopher, the founder of Gestalt theory, Max Wertheimer and the German-U.S. psychologists Kurt Koffka and Wolfgang Köhler, rejected the earlier assumption that perceptual organization was the product of learned relationships (associations), the constituent elements of which were called simple sensations. Although Gestaltists agreed that simple sensations logically could be understood to comprise organized percepts, they argued that percepts themselves were basic to experience. One does not perceive so many discrete dots (as simple sensations), for example; the percept is that of a dotted line.

Without denying that learning can play some role in perception, many theorists took the position that perceptual organization reflects innate properties of the brain itself. Indeed, perception and brain functions were held by Gestaltists to be formally identical (or isomorphic), so much so that to study perception is to study the brain. Much contemporary research in perception is directed toward inferring specific features of brain function from such behavior as the reports (introspections) people give of their sensory experiences. More and more such inferences are gratifyingly being matched with physiological observations of the brain itself.

Many investigators relied heavily on introspective reports, treating them as though they were objective descriptions of public events. Serious doubts were raised in the 1920s about this use of introspection by the U.S. psychologist John B. Watson and others, who argued that it yielded only subjective accounts and that percepts are inevitably private experiences and lack the objectivity commonly required of scientific disciplines. In response to objections about subjectivism, there arose an approach known as behaviorism that restricts its data to objective descriptions or measurements of the overt behavior of organisms other than the experimenter himself.

The fundamental principle of gestalt perception is the law of *prägnanz* (German for *pithiness*) which says that we tend to order our experience in a manner that is regular, orderly, symmetric, and simple. Gestalt psychologists attempt to discover refinements of the law of *prägnanz*, and this involves writing down laws which hypothetically allow us to predict the interpretation of sensation, what are often called "gestalt laws". These include:

Law of Closure — The mind may experience elements it does not perceive through sensation, in order to complete a regular figure (that is, to increase regularity).

Law of Similarity — The mind groups similar elements into collective entities or totalities.

This similarity might depend on relationships of form, color, size, or brightness.

Law of Proximity — Spatial or temporal proximity of elements may induce the mind to perceive a collective or totality.

Law of Symmetry (Figure ground relationships)— Symmetrical images are perceived collectively, even in spite of distance.

Law of Continuity — The mind continues visual, auditory, and kinetic patterns.

Law of Common Fate — Elements with the same moving direction are perceived as a collective or unit.

Not only does perception involve organization and grouping, it also involves distinguishing an object from its surroundings. Notice that once you perceive an object, the area around that object becomes the background. Gestalt psychologists have devised ambiguous figure-ground relationships - that is, drawings in which the figure and ground can be reversed - to illustrate their point that the whole is different from the sum of its parts.

The "figure and ground" illusion is commonly experienced when one gazes at the illustration of a black vase the outline of which is created by two white profiles. At any moment one will be able to see either the black vase (in the centre area) as "figure" or the white profiles on each side (in which case the black is seen as "ground"). The fluctuations of figure and ground may occur even when one fails deliberately to shift attention, appearing without conscious effort. Seeing one aspect apparently excludes seeing the other.

Although such illustrations may fool our visual systems, people are rarely confused about what they see. In real world, vases do not change into faces as we look at them. Instead, our perceptions are remarkably stable.

### PERSONALITY DISORDERS

Personality refers to a distinctive set of traits, behavior styles, and patterns that make up our character or individuality. How we perceive the world, our attitudes, thoughts, and feelings are all part of our personality. People with healthy personalities are able to cope with normal stresses and have no trouble forming relationships with family, friends, and co-workers. Those who struggle with a personality disorder have great difficulty dealing with other people. They tend to be inflexible, rigid, and unable to respond to the changes and demands of life. Although they feel that their behavior patterns are "normal" or "right," people with personality disorders tend to have a narrow view of the world and find it difficult to participate in social activities. A personality disorder must fulfill several criteria. A deeply ingrained, inflexible pattern of

relating, perceiving, and thinking serious enough to cause distress or impaired functioning is a personality disorder. Personality disorders are usually recognizable by adolescence or earlier, continue throughout adulthood, and become less obvious throughout middle age. Some experts believe that events occurring in early childhood exert a powerful influence upon behavior later in life. Others indicate that people are genetically predisposed to personality disorders. In some cases, however, environmental facts may cause a person who is already genetically vulnerable to develop a personality disorder. There are many formally identified personality disorders, each with their own set of behaviors and symptoms. Many of these fall into three different categories or clusters: Cluster A refers to odd or eccentric behavior; Cluster B refers to dramatic, emotional or erratic behavior; Cluster C refers to anxious fearful behavior. Since there are too many identified types of personality disorders to explain in this context, we will only review a few in each cluster. Cluster A: Schizoid Personality Disorder is a type of cluster A Schizoid personalities are introverted, withdrawn, solitary, emotionally cold, and distant. They are often absorbed with their own thoughts and feelings and are fearful of closeness and intimacy with others. For example, a person suffering from schizoid personality is more of a daydreamer than a practical action taker. Another example of cluster A is Paranoid Personality Disorder. The essential feature for this type of personality disorder is interpreting the actions of others as deliberately threatening or demeaning. People with paranoid personality disorder are untrusting, unforgiving, and prone to angry or aggressive outbursts without justification because they perceive others as unfaithful, disloyal, condescending or deceitful. This type of person may also be jealous, guarded, secretive, and scheming, and may appear to be emotionally "cold" or excessively serious. Antisocial Personality Disorder is an example of cluster B. People with antisocial personality disorder characteristically act out their conflicts and ignore normal rules of social behavior. These individuals are impulsive, irresponsible, and callous. Typically, the antisocial personality has a history of legal difficulties, belligerent and irresponsible behavior, aggressive and even violent relationships. They show no respect for other people and feel no remorse about the effects of their behavior on others. These people are at high risk for substance abuse, especially alcoholism, since it helps them to relieve tension, irritability and boredom. Avoidant Personality Disorder is an example of cluster C. Avoidant personalities are often hypersensitive to rejection and are unwilling to become involved with others unless they are sure of being liked. Excessive social discomfort, timidity, fear of criticism, avoidance of

social or work activities that involve interpersonal contact are characteristic of the avoidant personality. They are fearful of saying something considered foolish by others; worry they will blush or cry in front of others; and are very hurt by any disapproval by others. People with avoidant personality disorder may have no close relationships outside of their family circle, although they would like to, and are upset at their inability to relate well to others.

There are many types of help available for the different personality disorders. Treatment may include individual, group, or family psychotherapy. Medications, prescribed by a patient's physician, may also be helpful in relieving some of the symptoms of personality disorders, including problems with anxiety and perceptions.

Psychotherapy for patients with personality disorders focuses on helping them see the unconscious conflicts that are contributing to or causing their symptoms. It also helps people become more flexible and is aimed at reducing the behavior patterns that interfere with everyday living.

In psychotherapy, people with personality disorders can better recognize the effects of their behavior on others. Behavior and cognitive therapies focus on resolving symptoms or traits that are characteristic of the disorder, such as the inability to make important life decisions or the inability to initiate relationships.

## **Appendix B**

### **The Test Forms Used In Experiment 1**

#### **Form A**

##### **Questions (Attribution Theory)**

1. People tend to make external attributions for a behavior when that behavior is determined by
  - e. personal dispositions, traits, abilities and feelings
  - f. factors that the individual has no control over
  - g. situational demands or environmental factors
  - h. factors that the individual is able to control
2. People often do what they do, according to the attribution theory, because:
  - a. people want to feel more confident about themselves
  - b. people want to feel good about themselves
  - c. people want to see the world as simply as possible – what makes the most sense
  - d. people want to seem as intelligent as possible
3. According to the attribution theory, some factors affecting behavior are unstable. Which of the following factors is unstable?
  - a. Personality characteristics
  - b. ability
  - c. luck
  - d. task difficulty
4. How well a person believes he or she has performed determines:
  - a. how much she likes a given subject
  - b. how much time she will spend working on a project
  - c. how well she will do on an exam
  - d. how hard she will work on problem sets
5. The attribution theory provides explanations for many daily life phenomena. Which of the following phenomena can be explained by the attribution theory:
  - a. academic success
  - b. academic difficulty
  - c. academic motivation
  - d. academic stress
6. Factors such as ability and effort are considered to be:
  - a. stable factors
  - b. internal factors
  - c. controllable factors
  - d. none of these

7. Which of the following does not motivate people to continue doing what they were previously doing?
  - a. "I studied for 8 hours yesterday!"
  - b. "I'm never gone when she gives quizzes, and she never gives quizzes when I'm gone!"
  - c. "This test is so hard!"
  - d. "I love European history; it is so interesting!"
8. According to the attribution theory, in which of the following cases, the student will drop or fail the class?
  - a. Shirley was valedictorian of her high school class easily. She never studied for her classes, but score really well. Her freshman composition course has a weekly quiz over the previous week's material, for which she has to review for 15 minutes each night.
  - b. Harry reads his psychology book every night, but hates taking the daily quizzes given over the reading.
  - c. Jack always found school hard. He had to study for hours just to pull of B's in his classes. Now in college, his organic chemistry course requires him to study for three hours each night.
  - d. Mike never studies, but fortunately he always listens to just the right lectures.
9. The attribution theory emphasizes which of the following?
  - a. the difficulty of a subject
  - b. the amount of time an assignment takes
  - c. how lucky a person thinks she is
  - d. how well a person performs in a subject
10. Which factor of decision-making mentioned below encompasses the others?
  - a. I'm not doing well in physics because my professor doesn't explain things clearly.
  - b. I'm not doing well in physics, but if I study more, I know I'll pass.
  - c. I'm not doing well in physics, but that's because I'm more of an English person.
  - d. I'm not doing well in physics, but when I read the book, I do better on the quizzes.

### Questions (Personality Theory)

1. Which of the followings does not make up personality, according to the article?
  - a. feelings
  - b. patterns of behavior
  - c. outlook on life
  - d. instincts
2. Psychotherapy focuses on:
  - a. loosening people's tightly controlled emotions
  - b. helping people become more flexible
  - c. helping people control their anger

- d. helping people realize that rejection is not the ultimate loss
3. Antisocial Personality Disorder is associated with
    - a. avoidance of social or work activities
    - b. fearful of closeness
    - c. high risk of alcoholism
    - d. emotional coldness
  4. Someone who is emotionally withdrawn, distant, or solitary likely:
    - a. has Paranoid Personality Disorder
    - b. has an increased risk of substance abuse
    - c. has Schizoid Personality Disorder
    - d. has a cluster B disorder
  5. People with personality disorders find it most difficult to:
    - a. make decisions about things
    - b. follow simple instructions
    - c. behave appropriately in everyday situations
    - d. dealing with other people
  6. Cluster B refers to:
    - a. dramatic, emotional, or erratic behavior
    - b. odd and eccentric behavior
    - c. fearful and anxious behavior
    - d. passive and stoic behavior
  7. Currently, treatment for personality disorders does not include:
    - a. family psychotherapy
    - b. medication
    - c. occupation-based therapy
    - d. cognitive therapy
  8. Personality disorders have NOT been postulated to:
    - a. be caused by heritable factors
    - b. be caused by events occurring in childhood
    - c. increase in severity with age
    - d. be influenced by environmental factors
  9. In order for something to be considered a personality disorder, the following must be observed:
    - a. an inflexible way of relating to people
    - b. rapid changes in temperament
    - c. a pattern of deep thinking
    - d. difficulty in comprehending the world
  10. Someone with a personality disorder is most similar to which of the following in regards to perception of their own behavior?



- a. a middle aged man
- b. a five-year old girl
- c. a teenager
- d. a grandmother

### **Form B**

#### Questions (Attribution Theory)

1. An external attribution occurs when individuals attribute the cause of another person's behavior to
  - i. personal dispositions, traits, abilities and feelings
  - j. factors that the individual has no control over
  - k. situational demands or environmental factors
  - l. factors that the individual is able to control
2. The primary motivation behind people's worldviews according to the attribution theory:
  - a. people want to feel more confident about themselves
  - b. people want to feel good about themselves
  - c. people want to see the world as simply as possible – what makes the most sense
  - d. people want to seem as intelligent as possible
3. Which of the following factors is unstable according to the attribution theory?
  - a. Personality characteristics
  - b. ability
  - c. luck
  - d. task difficulty
4. A person's perception of success or failure determine:
  - a. how much she likes a given subject
  - b. how much time she will spend working on a project
  - c. how well she will do on an exam
  - d. how hard she will work on problem sets
5. The attribution theory explains implications of:
  - a. academic success
  - b. academic difficulty
  - c. academic motivation
  - d. academic stress
6. Ability and effort are both:
  - a. stable factors
  - b. internal factors
  - c. controllable factors
  - d. none of these
7. Which of the following does not influence people's motivation, according to the attribution theory?

- a. "I studied for 8 hours yesterday!"
  - b. "I'm never gone when she gives quizzes, and she never gives quizzes when I'm gone!"
  - c. "This test is so hard!"
  - d. "I love European history; it is so interesting!"
8. Which student is most likely to drop a course?
- a. Shirley was valedictorian of her high school class easily. She never studied for her classes, but score really well. Her freshman composition course has a weekly quiz over the previous week's material, for which she has to review for 15 minutes each night.
  - b. Harry reads his psychology book every night, and he is always ready for the daily quizzes given over the reading.
  - c. Jack always found school hard. He had to study for hours just to pull of B's in his classes. Now in college, his organic chemistry course requires him to study for three hours each night.
  - d. Mike never studies, but fortunately he always listens to just the right lectures.
9. Which of the following is most important according to the attribution theory?
- a. the difficulty of a subject
  - b. the amount of time an assignment takes
  - c. how lucky a person thinks she is
  - d. how well a person performs in a subject
10. Which is the most over-arching factor in decision-making according to the attribution theory?
- a. I'm not doing well in physics because my professor doesn't explain things clearly.
  - b. I'm not doing well in physics, but if I study more, I know I'll pass.
  - c. I'm not doing well in physics, but that's because I'm more of an English person.
  - d. I'm not doing well in physics, but when I read the book, I do better on the quizzes.

### Questions (Personality Theory)

1. According to the article, personality is made up of all of the followings except?
  - a. feelings
  - b. patterns of behavior
  - c. outlook on life
  - d. instincts
2. The basic idea behind psychotherapy is
  - a. loosening people's tightly controlled emotions
  - b. helping people become more flexible
  - c. helping people control their anger
  - d. helping people realize that rejection is not the ultimate loss
3. One of the symptoms of Antisocial Personality Disorder is

- a. avoidance of social or work activities
  - b. fearful of closeness
  - c. high risk of alcoholism
  - d. emotional coldness
4. Sarah tends to be emotionally withdrawn and she often feels distant from others. Sarah most likely:
- a. has Paranoid Personality Disorder
  - b. has an increased risk of substance abuse
  - c. has Schizoid Personality Disorder
  - d. has a cluster B disorder
5. Personality disorders cause otherwise normal people to find it most difficult to:
- a. make decisions about things
  - b. follow simple instructions
  - c. behave appropriately in everyday situations
  - d. dealing with other people
6. What are the symptoms associated with Cluster B
- a. dramatic, emotional, or erratic behavior
  - b. odd and eccentric behavior
  - c. fearful and anxious behavior
  - d. passive and stoic behavior
7. Which of the following is not treatment for personality disorders?
- a. family psychotherapy
  - b. medication
  - c. occupation-based therapy
  - d. cognitive therapy
8. Generally, personality disorders are not:
- a. caused by heritable factors
  - b. caused by events occurring in childhood
  - c. to increase in severity with age
  - d. influenced by environmental factors
9. Symptoms of a personality disorder include:
- a. an inflexible way of relating to people
  - b. rapid changes in temperament
  - c. a pattern of deep thinking
  - d. difficulty in comprehending the world
10. People with disordered personalities can have self-perception much like:
- a. a middle aged man
  - b. a five-year old girl
  - c. a teenager

- d. a grandmother

Questions  
(Gestalt Psychology)

1. Which of the following is NOT a Gestalt law?
  - a. Law of Symmetry
  - b. Law of Common Fate
  - c. Law of Distinction
  - d. Law of Closure
2. Gestalt theory aims to explain:
  - a. human decision-making
  - b. human perception organization
  - c. human reasoning
  - d. human perception processes
3. Gestalt psychology was founded on which of the following?
  - a. simple sensations are fundamental to organized perception
  - b. connections drawn from past experiences help organize perception
  - c. several laws guide the basic mechanism of perception
  - d. knowledge and perception are essentially the same entity, just in two different contexts
4. The Law of Common Fate refers to:
  - a. the tendency for people going to the same place to become friends
  - b. the perception of a dotted line to seem like a solid line when presented with a group of solid lines
  - c. the perception of balls moving in the same direction as a single mechanism
  - d. the tendency for people to view the world in the same light as others of their profession
5. The fact that most people tend to perceive visual elements that are close to each other as being a single group illustrates the Gestalt principle of
  - a. similarity
  - b. continuity
  - c. proximity
  - d. symmetry
6. Perceiving figures that contain gaps as complete figures illustrates the Gestalt principle of
  - a. closure
  - b. continuity
  - c. proximity
  - d. symmetry
7. The vase and two faces illusion is an example of which Gestalt law?
  - a. Law of Symmetry
  - b. Law of Continuity
  - c. Law of Closure
  - d. Law of Pragnanz
8. The issue of subjectivity in results is associated with which psychologist?

- a. Kurt Kofka
  - b. John Watson
  - c. Max Wertheimer
  - d. Alan Baddeley
9. Cognition does NOT include which of the following?
- a. remembering
  - b. knowing
  - c. feeling
  - d. thinking
10. "Make your essays logically structured and as succinct as possible for next Tuesday!"
- This assignment follows:
- a. the Law of Continuity
  - b. the Law of Common Fate
  - c. the Law of Proximity
  - d. the Law of Pragnanz

## Appendix C

### The Demographic Information Questionnaire used in Experiment 1, Experiment 2 and Experiment 4

Date: \_\_\_\_\_

#### Demographic Information

Below is a list of basic demographic questions. Please respond to the questions that you feel comfortable answering. You may skip any questions that you do not feel comfortable answering for any reason.

1. Gender: \_\_\_\_\_ Female \_\_\_\_\_

2. Age: \_\_\_\_\_ Year of Birth: \_\_\_\_\_

3. Major at Rice: \_\_\_\_\_

4. Choose all that apply:

Hispanic \_\_\_\_\_  
 African American \_\_\_\_\_  
 Caucasian \_\_\_\_\_  
 Native American \_\_\_\_\_  
 Asian-Pacific Islander \_\_\_\_\_  
 Other \_\_\_\_\_

5. Current GPA: \_\_\_\_\_

6. Are you familiar with the theories that you read about last week in the experiment?

Yes \_\_\_\_\_ No \_\_\_\_\_

7. If you responded "yes" to the previous question, please write down the theories that you are familiar with.

7. If you responded "yes" to the previous question, where did you learn about the theories?

8. If you responded "yes" to Question 6, please rate your level of knowledge of the theories.

|      |                  |              |      |        |
|------|------------------|--------------|------|--------|
| 1    | 2                | 3            | 4    | 5      |
| Poor | Not satisfactory | Satisfactory | Good | Expert |

## Appendix D

### Articles Used in Experiment 3

#### ATTRIBUTION THEORY

Attribution theory (Weiner, 1980, 1992) is probably the most influential contemporary theory with implications for academic motivation. It incorporates behavior modification in the sense that it emphasizes the idea that learners are strongly motivated by the pleasant outcome of being able to feel good about themselves. It incorporates cognitive theory and self-efficacy theory in the sense that it emphasizes that learners' current self-perceptions will strongly influence the ways in which they will interpret the success or failure of their current efforts and hence their future tendency to perform these same behaviors.

According to attribution theory, the explanations that people tend to make to explain success or failure can be analyzed in terms of three sets of characteristics: First, the cause of the success or failure may be internal or external. That is, we may succeed or fail because of factors that we believe have their origin within us or because of factors that originate in our environment. Second, the cause of the success or failure may be either stable or unstable. If the we believe cause is stable, then the outcome is likely to be the same if we perform the same behavior on another occasion. If it is unstable, the outcome is likely to be different on another occasion. Third, the cause of the success or failure may be either controllable or uncontrollable. A controllable factor is one which we believe we ourselves can alter if we wish to do so. An uncontrollable factor is one that we do not believe we can easily alter.

Note that this factor is distinct from the previous two categories. An internal factor can be controllable (we can control our effort by trying harder) or uncontrollable (most people cannot easily change their basic intellectual ability or change from being an introvert to being an extrovert). Likewise, an external factor can be controllable (a person failing a difficult course could succeed by taking an easier course) or uncontrollable (if calculus is difficult because it is abstract, it will still be abstract no matter what we do).

An important assumption of attribution theory is that people will interpret their environment in such a way as to maintain a positive self-image. That is, they will attribute their successes or failures to factors that will enable them to feel as good as possible about themselves. In general, this means that when learners succeed at an academic task, they are likely to want to

attribute this success to their own efforts or abilities; but when they fail, they will want to attribute their failure to factors over which they have no control, such as bad teaching or bad luck.

The basic principle of attribution theory as it applies to motivation is that a person's own perceptions or attributions for success or failure determine the amount of effort the person will expend on that activity in the future.

There are four factors related to attribution theory that influence motivation in education: ability, task difficulty, effort, and luck. In terms of the characteristics discussed previously, these four factors can be analyzed in the following way: Ability is a relatively internal and stable factor over which the learner does not exercise much direct control. Task difficulty is an external and stable factor that is largely beyond the learner's control. Effort is an internal and unstable factor over which the learner can exercise a great deal of control. Luck is an external and unstable factor over which the learner exercises very little control.

Note that it is the learner's perception that determines how attributions will influence future effort. A learner may believe that he is a "lucky person" - and for him luck would be an internal and stable characteristic over which he exercises little control. In other words, for this person "luck" is really what the preceding list calls an "ability" or personality characteristic. Likewise, a person may believe that she expended a great deal of effort, when in fact she did not, or that an objectively easy task was difficult. The basic principle of attribution theory as it applies to motivation is that a person's own perceptions or attributions for success or failure determine the amount of effort the person will expend on that activity in the future.

## GESTALT PSYCHOLOGY

Organizing raw sensory stimuli into meaningful experiences involves cognition, a set of mental activities that includes thinking, knowing, and remembering. Knowledge and experience are extremely important to perception, because they help us make sense of the input to our sensory systems.

How people perceive a well-organized pattern or whole, instead of many separate parts, is a topic of interest in Gestalt psychology. According to Gestalt psychologists, the whole is different than the sum of its parts. *Gestalt* is a German word meaning *configuration* or *pattern*.

A major goal of Gestalt theory in the 20th century was to specify the brain processes that might account for the organization of perception. Gestalt theorists, chief among them the



German-U.S. psychologist and philosopher, the founder of Gestalt theory, Max Wertheimer and the German-U.S. psychologists Kurt Koffka and Wolfgang Köhler, rejected the earlier assumption that perceptual organization was the product of learned relationships (associations), the constituent elements of which were called simple sensations. Although Gestaltists agreed that simple sensations logically could be understood to comprise organized percepts, they argued that percepts themselves were basic to experience. One does not perceive so many discrete dots (as simple sensations), for example; the percept is that of a dotted line.

Without denying that learning can play some role in perception, many theorists took the position that perceptual organization reflects innate properties of the brain itself. Indeed, perception and brain functions were held by Gestaltists to be formally identical (or isomorphic), so much so that to study perception is to study the brain. Much contemporary research in perception is directed toward inferring specific features of brain function from such behavior as the reports (introspections) people give of their sensory experiences. More and more such inferences are gratifyingly being matched with physiological observations of the brain itself.

Many investigators relied heavily on introspective reports, treating them as though they were objective descriptions of public events. Serious doubts were raised in the 1920s about this use of introspection by the U.S. psychologist John B. Watson and others, who argued that it yielded only subjective accounts and that percepts are inevitably private experiences and lack the objectivity commonly required of scientific disciplines. In response to objections about subjectivism, there arose an approach known as behaviorism that restricts its data to objective descriptions or measurements of the overt behavior of organisms other than the experimenter himself.

The fundamental principle of gestalt perception is the law of *prägnanz* (German for *pithiness*) which says that we tend to order our experience in a manner that is regular, orderly, symmetric, and simple. Gestalt psychologists attempt to discover refinements of the law of *prägnanz*, and this involves writing down laws which hypothetically allow us to predict the interpretation of sensation, what are often called "gestalt laws". These include:

Law of Closure — The mind may experience elements it does not perceive through sensation, in order to complete a regular figure (that is, to increase regularity).

Law of Similarity — The mind groups similar elements into collective entities or totalities.

This similarity might depend on relationships of form, color, size, or brightness.

Law of Proximity — Spatial or temporal proximity of elements may induce the mind to perceive a collective or totality.

Law of Symmetry (Figure ground relationships)— Symmetrical images are perceived collectively, even in spite of distance.

Law of Continuity — The mind continues visual, auditory, and kinetic patterns.

Law of Common Fate — Elements with the same moving direction are perceived as a collective or unit.

Not only does perception involve organization and grouping, it also involves distinguishing an object from its surroundings. Notice that once you perceive an object, the area around that object becomes the background. Gestalt psychologists have devised ambiguous figure-ground relationships - that is, drawings in which the figure and ground can be reversed - to illustrate their point that the whole is different from the sum of its parts.

The "figure and ground" illusion is commonly experienced when one gazes at the illustration of a black vase the outline of which is created by two white profiles. At any moment one will be able to see either the black vase (in the centre area) as "figure" or the white profiles on each side (in which case the black is seen as "ground"). The fluctuations of figure and ground may occur even when one fails deliberately to shift attention, appearing without conscious effort. Seeing one aspect apparently excludes seeing the other.

Although such illustrations may fool our visual systems, people are rarely confused about what they see. In real world, vases do not change into faces as we look at them. Instead, our perceptions are remarkably stable.

### PERSONALITY DISORDERS

Personality refers to a distinctive set of traits, behavior styles, and patterns that make up our character or individuality. How we perceive the world, our attitudes, thoughts, and feelings are all part of our personality. People with healthy personalities are able to cope with normal stresses and have no trouble forming relationships with family, friends, and co-workers. Those who struggle with a personality disorder have great difficulty dealing with other people. They tend to be inflexible, rigid, and unable to respond to the changes and demands of life. Although they feel that their behavior patterns are "normal" or "right," people with personality disorders tend to have a narrow view of the world and find it difficult to participate in social activities. A personality disorder must fulfill several criteria. A deeply ingrained, inflexible pattern of

relating, perceiving, and thinking serious enough to cause distress or impaired functioning is a personality disorder. Personality disorders are usually recognizable by adolescence or earlier, continue throughout adulthood, and become less obvious throughout middle age. Some experts believe that events occurring in early childhood exert a powerful influence upon behavior later in life. Others indicate that people are genetically predisposed to personality disorders. In some cases, however, environmental facts may cause a person who is already genetically vulnerable to develop a personality disorder. There are many formally identified personality disorders, each with their own set of behaviors and symptoms. Many of these fall into three different categories or clusters: Cluster A refers to odd or eccentric behavior; Cluster B refers to dramatic, emotional or erratic behavior; Cluster C refers to anxious fearful behavior. Since there are too many identified types of personality disorders to explain in this context, we will only review a few in each cluster. Cluster A: Schizoid Personality Disorder is a type of cluster A Schizoid personalities are introverted, withdrawn, solitary, emotionally cold, and distant. They are often absorbed with their own thoughts and feelings and are fearful of closeness and intimacy with others. For example, a person suffering from schizoid personality is more of a daydreamer than a practical action taker. Another example of cluster A is Paranoid Personality Disorder. The essential feature for this type of personality disorder is interpreting the actions of others as deliberately threatening or demeaning. People with paranoid personality disorder are untrusting, unforgiving, and prone to angry or aggressive outbursts without justification because they perceive others as unfaithful, disloyal, condescending or deceitful. This type of person may also be jealous, guarded, secretive, and scheming, and may appear to be emotionally “cold” or excessively serious. Antisocial Personality Disorder is an example of cluster B. People with antisocial personality disorder characteristically act out their conflicts and ignore normal rules of social behavior. These individuals are impulsive, irresponsible, and callous. Typically, the antisocial personality has a history of legal difficulties, belligerent and irresponsible behavior, aggressive and even violent relationships. They show no respect for other people and feel no remorse about the effects of their behavior on others. These people are at high risk for substance abuse, especially alcoholism, since it helps them to relieve tension, irritability and boredom. Avoidant Personality Disorder is an example of cluster C. Avoidant personalities are often hypersensitive to rejection and are unwilling to become involved with others unless they are sure of being liked. Excessive social discomfort, timidity, fear of criticism, avoidance of

social or work activities that involve interpersonal contact are characteristic of the avoidant personality. They are fearful of saying something considered foolish by others; worry they will blush or cry in front of others; and are very hurt by any disapproval by others. People with avoidant personality disorder may have no close relationships outside of their family circle, although they would like to, and are upset at their inability to relate well to others.

There are many types of help available for the different personality disorders. Treatment may include individual, group, or family psychotherapy. Medications, prescribed by a patient's physician, may also be helpful in relieving some of the symptoms of personality disorders, including problems with anxiety and perceptions.

Psychotherapy for patients with personality disorders focuses on helping them see the unconscious conflicts that are contributing to or causing their symptoms. It also helps people become more flexible and is aimed at reducing the behavior patterns that interfere with everyday living.

In psychotherapy, people with personality disorders can better recognize the effects of their behavior on others. Behavior and cognitive therapies focus on resolving symptoms or traits that are characteristic of the disorder, such as the inability to make important life decisions or the inability to initiate relationships.

### FREUD'S PERSONALITY THEORY

Freud did not exactly invent the idea of the conscious versus unconscious mind, but he certainly was responsible for making it popular. The conscious mind is what you are aware of at any particular moment, your present perceptions, memories, thoughts, fantasies, feelings, what have you. Working closely with the conscious mind is what Freud called the preconscious, what we might today call "available memory:" anything that can easily be made conscious, the memories you are not at the moment thinking about but can readily bring to mind. The largest part by far is the unconscious. It includes all the things that are not easily available to awareness, including many things that have their origins there, such as our drives or instincts. According to Freud, the unconscious is the source of our motivations, whether they be simple desires for food or sex, neurotic compulsions, or the motives of an artist or scientist. And yet, we are often driven to deny or resist becoming conscious of these motives, and they are often available to us only in disguised form.

Freudian psychological reality begins with the world, full of objects. Among them is a very special object, the organism. The organism is special in that it acts to survive and reproduce, and it is guided toward those ends by its needs -- hunger, thirst, the avoidance of pain, and sex.

A part of the organism is the nervous system, which has as one of its characteristics a sensitivity to the organism's needs. The nervous system, as id, translates the organism's needs into motivational forces called. The id works in keeping with the pleasure principle, which can be understood as a demand to take care of needs immediately. Just picture the hungry infant, screaming itself blue. It doesn't "know" what it wants in any adult sense; it just knows that it wants it and it wants it now. Unfortunately, although a wish for food, such as the image of a juicy steak, might be enough to satisfy the id, it isn't enough to satisfy the organism. The need only gets stronger, and the wishes just keep coming. You may have noticed that, when you haven't satisfied some need, such as the need for food, it begins to demand more and more of your attention, until there comes a point where you can't think of anything else. This is the wish or drive breaking into consciousness. Luckily for the organism, there is that small portion of the mind we discussed before, the conscious, that is hooked up to the world through the senses. Around this little bit of consciousness, during the first year of a child's life, some of the "it" becomes "I," some of the id becomes ego. The ego relates the organism to reality by means of its consciousness, and it searches for objects to satisfy the wishes that id creates to represent the organisms' needs. The ego, unlike the id, functions according to the reality principle, which says "take care of a need as soon as an appropriate object is found." It represents reality and, to a considerable extent, reason. However, as the ego struggles to keep the id (and, ultimately, the organism) happy, it meets with obstacles in the world. It occasionally meets with objects that actually assist it in attaining its goals. And it keeps a record of these obstacles and aides. In particular, it keeps track of the rewards and punishments meted out by two of the most influential objects in the world of the child -- mom and dad. This record of things to avoid and strategies to take becomes the superego. It is not completed until about seven years of age. There are two aspects to the superego: One is the conscience, which is an internalization of punishments and warnings. The other is called the ego ideal. It derives from rewards and positive models presented to the child. The conscience and ego ideal communicate their requirements to the ego with feelings like pride, shame, and guilt.

The ego deals with the demands of reality, the id, and the superego as best as it can. But when the anxiety becomes overwhelming, the ego must defend itself. It does so by unconsciously blocking the impulses or distorting them into a more acceptable, less threatening form. The techniques are called the ego defense mechanisms. Some of them are:

Denial involves blocking external events from awareness. If some situation is just too much to handle, the person just refuses to experience it. As you might imagine, this is a primitive and dangerous defense -- no one disregards reality and gets away with it for long.

Repression is just that: not being able to recall a threatening situation, person, or event. This, too, is dangerous, and is a part of most other defenses.

Displacement is the redirection of an impulse onto a substitute target. If the impulse, the desire, is okay with you, but the person you direct that desire towards is too threatening, you can displace to someone or something that can serve as a symbolic substitute.

Projection is almost the complete opposite of turning against the self. It involves the tendency to see your own unacceptable desires in other people. In other words, the desires are still there, but they're not your desires anymore. I confess that whenever I hear someone going on and on about how aggressive everybody is, or how perverted they all are, I tend to wonder if this person doesn't have an aggressive or sexual streak in themselves that they'd rather not acknowledge.

### THE SEVEN SINS OF MEMORY

Despite memory's obvious benefits, it can also let us down, said Daniel Schacter, PhD, longtime memory researcher and chair of Harvard University's psychology department, at an APA 2003 Annual Convention session honoring the publication of his book, "The Seven Sins of Memory: How the Mind Forgets and Remembers" (Houghton Mifflin, 2001).

"Memory, for all that it does for us every day...for all the feats that can sometimes amaze us, can also be a troublemaker," said Schacter of his book, which describes the seven major categories of memory foibles being investigated by psychologists.

However, noted Schacter, the same brain mechanisms account for memory's sins as well as its strengths, so investigating its negatives exposes its positives. "We shouldn't think of these fundamentally as flaws in the architecture of memory," he explained, "but rather as costs we pay for benefits in memory that make it work as well as it does most of the time."

At the session, during which Schacter received the APA Div. 1 (Society for General Psychology) William James Book Award, he defined his book's seven sins. The first three are "sins of omission" that involve forgetting, and the second four are "sins of commission" that involve distorted or unwanted recollections.

Transience--the decreasing accessibility of memory over time. While a degree of this is normal with aging, decay of or damage to the hippocampus and temporal lobe can cause extreme forms of it. Schacter cited as a somewhat facetious example former President Bill Clinton's "convenient lapses of memory" during the Monica Lewinsky investigation. Clinton claimed in the hearings that he sometimes couldn't remember what had happened the previous week.

Absent-mindedness--lapses of attention and forgetting to do things. This sin operates both when a memory is formed (the encoding stage) and when a memory is accessed (the retrieval stage). Examples, said Schacter, are forgetting where you put your keys or glasses. He noted a particularly famous instance in which cellist Yo-Yo Ma forgot to retrieve his \$2.5 million cello from the trunk of a New York City cab.

Blocking--temporary inaccessibility of stored information, such as tip-of-the-tongue syndrome. Schacter recounted the embarrassment of John Prescott, British deputy prime minister, when a reporter asked him how the government was paying for the expensive Millennium Dome. Prescott struggled to find the word "lottery," trying "raffles" instead.

Suggestibility--incorporation of misinformation into memory due to leading questions, deception and other causes. Psychologists Elizabeth Loftus, PhD, and Stephen Ceci, PhD, are among those well-known in this research (see sidebar).

Bias--retrospective distortions produced by current knowledge and beliefs. Psychologist Michael Ross, PhD, and others have shown that present knowledge, beliefs and feelings skew our memory for past events, said Schacter. For example, research indicates that people currently displeased with a romantic relationship tend to have a disproportionately negative take on past states of the relationship.

Persistence--unwanted recollections that people can't forget, such as the unrelenting, intrusive memories of post-traumatic stress disorder. An example, said Schacter, is the case of Donnie Moore of the California Angels, who threw the pitch that lost his team the 1986 American League Championship against the Boston Red Sox. Moore fixated on the bad play, said Schacter, "became a tragic prisoner of memory," and eventually committed suicide.

Misattribution--attribution of memories to incorrect sources or believing that you have seen or heard something you haven't. Prominent researchers in this area include Henry L. Roediger III, PhD, and Kathleen McDermott, PhD. An illustration of it, said Schacter, is the rental shop mechanic who thought that an accomplice, known as "John Doe No. 2," had worked with Timothy McVeigh in the Oklahoma City bombing; he thought he'd seen the two of them together in his shop. In fact, the mechanic had encountered John Doe No. 2 alone on a different day.

Schacter has focused on this last area in his own research. He's been probing the neuropsychology of why people "misremember" having seen words. His work with amnesiacs and normal participants indicates that people's normal tendency to remember "the gist of a list" of semantically similar words--a tendency missing in amnesiacs--is also what causes them to misremember words not on the list.

In his latest line of research, Schacter is using imaging to detect the brain mechanisms at work in false and correct recognition of words and shapes--work, he said, which "highlights that by using cognitive neuroscience, we can start to home in on some of the brain mechanisms involved in each of the sins."

He added, "Ultimately we think this research will help us to establish a unified view of these seven sins of memory."



## Appendix E

### Factual and Conceptual Questions Used in Experiment 2

#### ATTRIBUTION THEORY

##### Conceptual Questions

1. Belle watches Jose slip on the stairs of their apartment building. If Bella decides that Jose is an uncoordinated person who was not paying attention to what he was doing, she has made
  - e. an external attribution
  - f. a self-serving attribution
  - g. an attributional error
  - h. an internal attribution
2. Tina recently lost over half of the money she had put away for her retirement. Her close friend Diane concludes that Tina lost the money because the stock market took a significant downturn, she has made
  - a. an internal attribution
  - b. the fundamental attribution error
  - c. an illusionary correlation
  - d. an external attribution
3. Imagine that you recently applied for a job, but the job went to someone else. According to the attribution theory, if you state: "It was bad luck that the boss had a neighbor who applied the same day that I did," you have made
  - a. an internal-unstable attribution
  - b. an external-stable attribution
  - c. an external-unstable attribution
  - d. an internal-stable attribution
4. The following statement emphasizes the importance of \_\_\_\_\_ factor that influences motivation in education: "I'm never gone when she gives quizzes, and she never gives quizzes when I'm gone!"
  - a. ability
  - b. task difficulty
  - c. effort
  - d. luck
5. Parker is usually extremely well prepared for the exams but yesterday he could not study for the exam and he got a D. for the next exam, Parker is determined to study hard. Parker is trying to control which one of the factors related to attribution theory that influence motivation in education?
  - a. ability
  - b. task difficulty
  - c. effort
  - d. luck

##### Factual Questions

1. When a(n) \_\_\_\_\_ is made, the cause of the given behavior is assigned to the individual's personality, attitudes, character or disposition.
  - e. an external attribution
  - f. a self-serving attribution
  - g. an attributional error
  - h. an internal attribution
  
2. When a(n) \_\_\_\_\_ is made, the cause of the given behavior is assigned to the situation in which the behavior was seen.
  - a. an internal attribution
  - b. the fundamental attribution error
  - c. an illusory correlation
  - d. an external attribution
  
3. \_\_\_\_\_ occurs when we attribute the cause of the success or failure to environmental factors. On the other hand \_\_\_\_\_ occurs when we believe we ourselves can change the situation if we wish to do so.
  - a. an internal-unstable attribution
  - b. an external-stable attribution
  - c. an external-unstable attribution
  - d. an internal-stable attribution
  
4. \_\_\_\_\_ is an external and unstable factor over which the student has very little control.
  - a. ability
  - b. task difficulty
  - c. effort
  - d. luck
  
5. \_\_\_\_\_ is an internal and unstable factor which influences motivation in education.
  - a. ability
  - b. task difficulty
  - c. effort
  - d. luck

## **GESTALT PSYCHOLOGY**

### **Conceptual Questions**

1. Ray sat on his porch looking out at his field of corn. The fact that Ray perceived the corn in the field as being grouped into a series of separate rows is consistent with the Gestalt principle of
  - a. similarity
  - b. continuity
  - c. proximity
  - d. symmetry

2. When Wade looked up at the sky he perceived the seven stars that made up the Big Dipper as a single complete figure, rather than as individual stars. Wade's perception of the night sky illustrates the Gestalt principle of
  - a. closure
  - b. continuity
  - c. proximity
  - d. symmetry
3. Sarah held a small earthworm in each hand and then held her hands together so that only one end of each of the earthworms could be seen. In this way, she was able to fool her little brother into thinking she had one gigantic earthworm in her hands. Sarah's trick illustrates the Gestalt principle of
  - a. similarity
  - b. symmetry
  - c. closure
  - d. continuity
4. "Make your essays logically structured and as succinct as possible for next Tuesday!" This assignment follows:
  - a. Law of closure
  - b. Law of good from
  - c. Law of pragnanz
  - d. Law of continuity
5. Gwen was at a basketball game, and even though people wearing red shirts were spread evenly through the stands, she perceived all the red shirts as a single group of visiting fans. Gwen's perception is most consistent with the Gestalt principle of
  - a. proximity
  - b. continuity
  - c. closure
  - d. similarity

### **Factual Questions**

11. The fact that most people tend to perceive visual elements that are close to each other as being a single group illustrates the Gestalt principle of
  - a. similarity
  - b. continuity
  - c. proximity
  - d. symmetry
12. Perceiving figures that contain gaps as complete figures illustrates the Gestalt principle of
  - a. closure
  - b. continuity
  - c. proximity
  - d. symmetry
13. Law of \_\_\_\_\_ states that human mind continues visual, auditory, and kinetic patterns.
  - a. similarity

- b. symmetry
  - c. closure
  - d. continuity
14. According to the article, \_\_\_\_\_ states that we tend to order our experience in a manner that is regular, orderly, symmetric, and simple.
- a. Law of closure
  - b. Law of good from
  - c. Law of pragnanz
  - d. Law of continuity
15. Most people tend to group similar elements into collective entities or totalities. This tendency illustrates the Gestalt principle of
- a. proximity
  - b. continuity
  - c. closure

## **FREUD'S PERSONALITY THEORY**

### **Conceptual Questions**

1. Donavon hates his boss. To deal with the anxiety of working for someone he dislikes intensely Donavon has convinced himself that his boss can't stand him, and that his boss purposely gives Donavon all the bad assignment in the company. According to Freud, Donavon may be unconsciously dealing with the anxiety caused by his feeling towards his boss by using the defense mechanisms of
  - a. reaction formation
  - b. denial
  - c. displacement
  - d. projection
2. Roland has extremely strong moral standards. However, whenever he fails to live up to the high standards he sets for himself, Roland becomes overwhelmed by feelings of guilt and anxiety. Freud would most likely suggest that Roland's personality is dominated by his
  - a. feelings of incongruence
  - b. superego
  - c. id
  - d. ego
3. When Jacob was five he was burned quite badly while he was playing in the kitchen. Jacob is now 30 years old, and he has no memory of the accident or the hospital stay. When his parents talk to him about it he is convinced that they are confused and that the accident must have happened to one of his brothers. According to Freud, the memory of accident is in Jacob's
  - a. preconscious
  - b. unconscious
  - c. conscious
  - d. subconscious

4. 4 year-old Sarah was watching a cartoon on TV when a commercial for a horror movie came on. Sarah's mother wasn't able to save her child from this horror commercial, so she tried to comfort Sarah: "That was a scary commercial, wasn't it?" Sarah said "What commercial?" According to Freud, Sarah is using the defense mechanism of
  - a. denial
  - b. repression
  - c. projection
  - d. displacement
5. Jane is frustrated by her superiors at work. Sometimes she goes home and kicks the dog, beats up a family member, or engages in cross-burnings. According to Freud, Jane is using the defense mechanism of
  - a. projection
  - b. denial
  - c. reaction formation
  - d. displacement

### **Factual Questions**

1. \_\_\_\_\_ involves the tendency to attribute your own repressed thoughts to someone else
  - a. reaction formation
  - b. denial
  - c. displacement
  - d. projection
2. \_\_\_\_\_ can be thought of as a type of conscience that punishes misbehavior with feelings of guilt.
  - a. feelings of incongruence
  - b. superego
  - c. id
  - d. ego
3. \_\_\_\_\_ mind might be defined as that part of the mind which gives rise to a collection of mental phenomena that manifest in a person's mind but which the person is not aware of at the time of their occurrence.
  - a. preconscious
  - b. unconscious
  - c. conscious
  - d. subconscious
4. \_\_\_\_\_ occurs when an individual tries to avoid something uncomfortable. It involves blocking external events from awareness.
  - a. denial
  - b. repression
  - c. projection
  - d. displacement
5. The defense mechanism that is at work when a person diverts emotional feelings from their original source to a substitute target is
  - a. projection

- b. denial
- c. reaction formation
- d. displacement

### SEVEN SINS OF MEMORY

#### **Conceptual Questions**

1. 70- year old Sam has been experiencing some memory problems recently. He forgets where he puts his reading glasses. What is the memory sin he is committing?
  - a. transience
  - b. sins of commission
  - c. blocking
  - d. absent-mindedness
2. Sarah confuses a dream for a memory. She is committing the sin of \_\_\_\_\_.
  - a. transience
  - b. omission
  - c. commission
  - d. bias
3. A person who witnesses a murder after watching a television program may incorrectly blame the murder on someone she saw on the television program. This error is known as
  - a. persistence
  - b. transience
  - c. misattribution
  - d. blocking
4. Alex cannot remember what he ate in the breakfast. According to Schacter, Alex is experiencing the sin of
  - a. blocking
  - b. transience
  - c. absent-mindedness
  - d. bias
5. A person sees a crime being committed by a redheaded man. After reading in the newspaper that the crime was committed by a brown-haired man, the witness "remembers" a brown-haired man instead of a redheaded man. According to Schacter, this person experiences the sin of
  - a. bias
  - b. suggestibility
  - c. absent-mindedness
  - d. transience

#### **Factual Questions**

1. Among seven sins of memory, \_\_\_\_\_ refers to memory flaws when an individual forgets to do things due to lapses of attention.
  - a. transience

- b. sins of commission
  - c. blocking
  - d. absent-mindedness
2. \_\_\_\_\_ involves unwanted memories which are often distorted version of reality.
- a. transience
  - b. omission
  - c. commission
  - d. bias
3. \_\_\_\_\_ entails correct recollection of information with incorrect recollection of the source of that information
- a. persistence
  - b. transience
  - c. misattribution
  - d. blocking
4. \_\_\_\_\_ refers to the general deterioration of a specific memory over time.
- a. blocking
  - b. transience
  - c. absent-mindedness
  - d. bias
5. \_\_\_\_\_ involves integration of misinformation into memory.
- a. bias
  - b. suggestibility
  - c. absent-mindedness
  - d. transience

## PERSONALITY THEORY

### Conceptual Questions

1. Jose is suffering from \_\_\_\_\_. He does nothing but daydream. He is often absorbed by his own thought and feelings.
- a. Paranoid Personality Disorder
  - b. Schizoid Personality Disorder
  - c. Avoidant Personality Disorder
  - d. Antisocial Personality Disorder
2. Janet was recently diagnosed with \_\_\_\_\_. She has serious problems with her husband. She is extremely jealous and secretive. She is untrusting, unforgiving and aggressive.
- a. Paranoid Personality Disorder
  - b. Schizoid Personality Disorder
  - c. Avoidant Personality Disorder
  - d. Antisocial Personality Disorder

3. Sam breaks the law repeatedly. He lies, steals, and fights often. He disregards the safety of self and others and does not show any guilt. Sam displays the symptoms of \_\_\_\_\_.
- Paranoid Personality Disorder
  - Schizoid Personality Disorder
  - Avoidant Personality Disorder
  - Antisocial Personality Disorder
4. George was diagnosed with \_\_\_\_\_. He considers himself to be personally unappealing, and avoids social interaction for fear of being ridiculed, humiliated, rejected, or disliked.
- Paranoid Personality Disorder
  - Schizoid Personality Disorder
  - Avoidant Personality Disorder
  - Antisocial Personality Disorder
5. Anne is struggling with a personality disorder. She has great difficulty with dealing with other people. She cannot respond to the demands of life. There are many types of help that she can receive. According to the passage, which of the following options is not one of them?
- Psychotherapy which helps people become more flexible
  - Behavior and cognitive therapies which focus on resolving symptoms
  - Medications which aim to resolve the symptoms
  - Individual psychotherapy which help people see the unconscious conflicts that cause the symptoms

### **Factual Questions**

- \_\_\_\_\_ is associated with introversion, social withdrawal, and distant behavior.
  - Paranoid Personality Disorder
  - Schizoid Personality Disorder
  - Avoidant Personality Disorder
  - Antisocial Personality Disorder
- The essential feature of \_\_\_\_\_ is interpretation of the actions of others as deliberately threatening or demeaning.
  - Paranoid Personality Disorder
  - Schizoid Personality Disorder
  - Avoidant Personality Disorder
  - Antisocial Personality Disorder
- People with \_\_\_\_\_ act out their conflicts and ignore the normal rules of social behavior.
  - Paranoid Personality Disorder



- b. Schizoid Personality Disorder
  - c. Avoidant Personality Disorder
  - d. Antisocial Personality Disorder
4. People with \_\_\_\_\_ are often hypersensitive to rejection and are unwilling to become involved with others unless they are sure of being liked.
- a. Paranoid Personality Disorder
  - b. Schizoid Personality Disorder
  - c. Avoidant Personality Disorder
  - d. Antisocial Personality Disorder
5. According to the passage, which of the following options is not a type of help available for the different personality disorders?
- a. Psychotherapy which helps people become more flexible
  - b. Behavior and cognitive therapies which focus on resolving symptoms
  - c. Medications which aim to resolve the symptoms
  - d. Individual psychotherapy which help people see the unconscious conflicts that cause the symptoms

## Appendix F

### The Questions and the Type of Questions in the “Practice” Test Used In Experiment 3.

1. A psychologist who thinks that environmental factors determine human behavior would probably be considered a
  - A. physiological psychologist.
  - B. behaviorist.
  - C. social psychologist.
  - D. cognitive psychologist.
2. A developmental psychologist is interested in
  - A. how environment shapes human behavior.
  - B. how childhood experiences affects personality development.
  - C. how behavior changes as a function of age.
  - D. mental processes of newborns such as memory and learning.
3. Anna and Mary are identical twins and they have been raised together. Their personalities and intelligence levels are comparable. However, recently, Anna was diagnosed with cancer. Although they are identical twins, Mary did not develop cancer. This example reflects the underlying theme that
  - A. genetic factors and experience interact to shape behavior.
  - B. our cultural backgrounds exert a considerable influence over our behavior.
  - C. motives and expectations can color our experiences.
  - D. theoretical diversity is an important component in science.
4. A hypothesis is
  - A. a prediction based on a theory or a model.
  - B. a collection of scientific statements.
  - C. is a way of inferring cause-and-effect relationship.
  - D. a variable effected by IV.
5. A dependent variable
  - A. is manipulated by the experimenter.
  - B. has an effect on the other variables in the experiment.
  - C. is utilized only in experimental research.
  - D. is effected by the manipulated variable.
6. An experiment should include a control group in order to
  - A. establish the validity of the conclusions.
  - B. establish the reliability of the conclusions.
  - C. to control the variability in the dependent variable.
  - D. to control the variability in the independent variable.
7. Random assignment ensures that
  - A. every individual in the population has equal probability of ending up in the subject sample.
  - B. each subject in the sample has an equal probability of ending up in any experimental group.
  - C. results of the experiment are valid.
  - D. the experiment is ecologically valid.

8. Ashley wants to examine whether caffeine causes insomnia. What is the most suitable research design to examine this hypothesis?
- Correlational
  - Quasi-experimental
  - Experimental
  - Longitudinal
9. A and B correlate 0.07. Which of the following statements is correct about the association between A and B?
- They are casually related.
  - You can predict B without error if you know A.
  - You cannot predict B even if you know A.
  - None of the above is correct.
10. Anna conducted an experiment to investigate the effect of anxiety on problem solving skills. She found out that anxiety and problem solving skills are significantly associated. In other words, she meant that,
- her results had practical significance.
  - her results were due to chance.
  - her results were not due to chance.
  - her results are important.
11. Tony has Parkinson disease. As the disease progresses, Tony will experience some changes in his brain and nervous system. Which of the followings would you expect to see in his central nervous system?
- a lack of neurotransmitters in some neurons
  - areas where the myelin sheath has degenerated
  - areas where the dendrites are severely damaged
  - a reduction in the number of chloride ions in his peripheral nervous system
12. Elena has developed a specific phobia, arachnophobia. She tends to feel uneasy in any area she believes could harbor spiders or that has visible signs of their presence, such as webs. Whenever she sees a spider, you expect to see a significant activity in her
- her amygdale.
  - cerebellum.
  - temporal lobe.
  - the midbrain.
13. Lisa's grandmother had a stroke on the left side of her brain. After the stroke, her body could not function as it should, The stroke produced all of the following except
- speech and language problems.
  - paralysis on the right side of her body.
  - visual and auditory problems.
  - slow behavior.
14. Sam and Jay are fraternal twins who have been raised together. Their intelligence scores tend to show similarity yet this similarity is smaller than between Jack and Tom who are identical twins and have been raised separate. This information suggest that
- environmental factors are more important than genetic factors in determining intelligence level.
  - neither genetic nor environmental factors contribute to intelligence.

- C. genetic factors are more important than environmental factors.
  - D. both genetic and environmental factors contribute equally to this disorder.
15. The correlation coefficient between variable A and variable B shows (that)
- A. how strongly A influences B.
  - B. the strength of relationship between A and B.
  - C. how well A can predict B.
  - D. none of the above.
16. The mean is
- A. the number separating higher half of a sample from the lower half.
  - B. the value that occurs most frequently in a data set.
  - C. measure of the variability or dispersion.
  - D. average score in a data set.
17. Hindsight bias is
- A. tendency to see events that have occurred as more predictable than they were before they took place.
  - B. A tendency towards a particular perspective or result.
  - C. the tendency to search for or interpret information in a way that confirms ones preconceptions.
  - D. Tendency to ignore available statistical data in favor of particulars.
18. One study found that height and shoe size are correlated. Which one of the following can be inferred from this study?
- A. Shoe size predicts height.
  - B. Height predicts shoe size.
  - C. Tall people tend to have bigger feet.
  - D. Short people tend to have bigger feet.
19. Structuralism is associated with
- A. Skinner
  - B. Pavlov
  - C. Watson
  - D. Wundt
20. Feeling of a hot cup of coffee is possible through \_\_\_\_\_.
- A. Medulla.
  - B. Cerebellum.
  - C. Sensory neurons.
  - D. Motor neurons.
21. Sara's cat has been having problems with her balance. She cannot move in a smooth and coordinated manner. It is most probable that she has suffered damage to her
- A. Amygdale.
  - B. Angular gyrus.
  - C. Cerebellum.
  - D. Corpus collosum.
22. Jack hears a very loud sound. Such a strong stimulus:
- A. increases the intensity of a neuron's action potential.
  - B. affects the speed that a neuron fires.
  - C. triggers more neurons to fire.
  - D. does not affect how often a neuron fires.

23. Psychoanalytic perspective focuses on
- A. the impact of parents' child-rearing practices on children's personality development.
  - B. the effect of genetic factors on personality.
  - C. the effect of self-actualization on personality development.
  - D. the effect of evolution on gender development.
24. Your four-year old child has developed an ability to infer other's mental states. What characteristics of the preoperational stage is he exhibiting?
- A. The child is exhibiting object permanence.
  - B. The child is exhibiting conservation of mass.
  - C. The child is exhibiting reversibility of thinking.
  - D. The child is exhibiting the increase in theory of mind.
25. Jackie's 4-year old son has been exhibiting some growth problems. He is underweight and shorter than his peers. His symptoms relate to some hormones that are missing in his blood. It is likely that he has problems in his
- A. Limbic system.
  - B. Thalamus.
  - C. Hypothalamus.
  - D. Hippocampus.
26. Two-year old Gabriella has learned the schema for cow from her picture books. When she saw a moose, she called it a cow. Gabriella's overgeneralization is an example of
- A. Assimilation.
  - B. Accommodation.
  - C. Centration.
  - D. Conservation.
27. Adapting our current understandings (schemas) to incorporate new information is called
- A. Assimilation.
  - B. Conservation.
  - C. Accommodation.
  - D. Object Permanence.
28. Seven-year old Jack started doing arithmetical operations. According to Piaget, Jack is in the
- A. Pre-operational period of cognitive development
  - B. Assimilation vs. Accommodation period of cognitive development
  - C. Formal operational period of cognitive development
  - D. Concrete operational period of cognitive development
29. Whenever he sees a stranger, Ben feels distressed and calls for his mother. Based on this information, Ben should be in
- A. Pre-operational period of cognitive development
  - B. Sensorimotor period of cognitive development
  - C. Formal operational period of cognitive development
  - D. Concrete operational period of cognitive development
30. A limited ability to share another person's point of view is called
- A. Theory of mind
  - B. Egocentrism
  - C. Stranger anxiety

- D. Empathy
31. Complying with rules to avoid punishment or to obtain rewards are indicators of
- A. pre-conventional level of moral reasoning
  - B. conventional level of moral reasoning
  - C. postconventional level of moral reasoning
  - D. none of the above
32. \_\_\_\_\_ controls breathing, heart rate and blood pressure.
- A. Medulla
  - B. Cerebellum
  - C. Hypothalamus
  - D. Reticular Formation
33. Humanistic psychology emphasized
- A. observing and recording people's behavior as they respond to different situations
  - B. brain activity linked with mental activity
  - C. the meaning of early childhood memories
  - D. importance of current environmental influences on our growth potential
34. The principle that among the range of inherited trait variations, those contributing to reproduction and survival will most likely to be passed on to succeeding generations is known as
- A. heredity
  - B. natural selection
  - C. evolution
  - D. genetics
35. The approach which focuses on the influences of biological, psychological and social-cultural factor on mental processes is known as
- A. Sociocultural approach
  - B. Psychological approach
  - C. Social-psychobiological approach
  - D. Bio-psychosocial approach
36. \_\_\_\_\_ builds psychology's knowledge whereas \_\_\_\_\_ tackles practical problems.
- A. research, industry
  - B. basic research, applied research
  - C. applied research, basic research
  - D. theoretical research, practical research
37. \_\_\_\_\_ is a branch of psychology that assists people with problems in living and achieving greater well-being
- A. Counseling
  - B. Clinical
  - C. Psychiatry
  - D. Therapy
38. \_\_\_\_\_ is an observation technique in which one person is studied in depth in the hope of revealing universal principle.
- A. Survey
  - B. Experiment
  - C. Case study

- D. Longitudinal study
39. The level of stimulation required to trigger a neural impulse is called
- A. reuptake
  - B. action potential
  - C. threshold
  - D. all-or-none
40. \_\_\_\_\_ cortex is the area at the front of the parietal lobes that registers and processes body touch and movement sensations.
- A. Motor
  - B. Sensory
  - C. Cerebral
  - D. Visual
41. The brain's ability to change, especially during childhood by reorganizing after damage or by building new pathways based on experience is called
- A. plasticity
  - B. psychomotor development
  - C. amnesia
  - D. reorganization
42. The hemispheric specialization is known as
- A. split brain
  - B. ipsilateralism
  - C. lateralization
  - D. neurogenesis
43. Jordan is experiencing a serious sleep disorder. He experiences extreme fatigue and falls asleep at inappropriate times, such as while at work or at school. It is most likely that Jordan is suffering from
- A. insomnia
  - B. narcolepsy
  - C. sleep apnea
  - D. cataplexy
44. Most parents will tell you that after having their second child, babies differ even before gulping their first breath. This difference in quickly apparent aspect of personality is a function of
- A. personality differences
  - B. temperament differences
  - C. genetic differences
  - D. cultural differences
45. The parental statement 'Big boys don't cry Alex' is associated with
- A. social learning theory
  - B. cognitive dissonance theory
  - C. cognitive theory
  - D. cultural differentiation theory
46. Kathy notices that her baby develops some skills over a night. For example, one day, the baby could not grasp a pen but next day she can do it effortlessly. Kathy's baby is demonstrating
- A. continual development

- B. stability
  - C. gradual development
  - D. staged development
47. Among Aka people of Central Africa, fathers form an especially close bonds with their infant, even suckling the babies with their own nipples when hunger makes the child impatient for Mother's return. This type of relationship between father and the baby makes the babe develop \_\_\_\_\_.
- A. secure attachment
  - B. basic trust
  - C. separation anxiety
  - D. self-concept
48. By varying the way they look, adolescents try out different selves. For Erikson the biggest issue during the adolescence concerns
- A. intimacy
  - B. isolation
  - C. identity
  - D. generativity
49. It has been found that Americans are unhappiest when alone and happiest when with friends. This finding supports Erikson's stage of
- A. adolescence
  - B. young adulthood
  - C. middle adulthood
  - D. late adulthood
50. Habitual tasks, such as remembering to take medications three times daily, can be challenging for older adults due to
- A. psychomotor slowing
  - B. failure in their prospective memory skills
  - C. lack of myelin sheath
  - D. lack of neurotransmitters
51. A study found that humans laugh 30 times more often in social situations than in solitary situations. It is most likely that this study used
- A. experimental design
  - B. correlational design
  - C. survey
  - D. naturalistic observation
52. A study revealed that 72% of Americans think there is too much TV violence. This study most likely used
- A. naturalistic observation
  - B. survey method
  - C. experimental design
  - D. longitudinal design
53. Staying up especially late on weekends is most likely to have an influence on:
- A. narcolepsy.
  - B. sleep apnea.
  - C. the circadian rhythm.



- D. seasonal affective disorder.
54. Court systems frequently ban testimony from witnesses who have been hypnotized because the procedure often encourages:
- A. a hidden observer.
  - B. hallucinations.
  - C. narcolepsy.
  - D. false memories.
55. Your friend insists that she rarely has dreams. What is the most likely explanation for this situation?
- A. She probably has dreams every night, but only rarely remembers her dreams.
  - B. She probably uses barbiturates to help her sleep every night.
  - C. This is normal; most people never dream.
  - D. She stays up too late at night and gets up too early in the morning.
56. Racial and ethnic stereotypes can sometimes bias our perceptions of others' behaviors. This best illustrates the impact of:
- A. retinal disparity.
  - B. interposition.
  - C. top-down processing.
  - D. perceptual adaptation.
57. As you talk with a friend at a party, her voice is distinct from all the other voices you hear. This illustrates the perceptual principle of:
- A. grouping.
  - B. proximity.
  - C. closure.
  - D. figure-ground.
58. Olivia has trouble staying asleep for more than four hours at a time. Olivia is likely to experience
- A. less deep sleep than people who sleep for a full 8 hours
  - B. less REM sleep than people who sleep for a full 8 hours
  - C. less REM sleep and less deep sleep than people who sleep for a full 8 hours
  - D. more sleep spindles during Stage 3 and 4 of the sleep cycle
59. Which of the following activities requires you to use the absolute threshold for sensation?
- A. Telling the difference between sweet and salty
  - B. Detecting a tiny, faint light on a radar screen
  - C. Deciding if two glasses contain the same amount of water
  - D. Telling if your guitar is in tune
60. As we perceive the environment, we use Gestalt principles of good form to help organize things into meaningful groups and forms. In observing a football game on TV we are immediately able to recognize the players as members of their respective teams, regardless of their positions on the field. Which Gestalt principle is most responsible for this ability?
- A. Closure
  - B. Proximity
  - C. Similarity
  - D. Continuity

61. You are an artist and you are interested in completing a painting that gives a three-dimensional appearance. You want to show a beach in the foreground of the painting, people swimming in the ocean, and sailboats in the distance. Which perceptual principles will be most helpful as you complete your painting?
- A. binocular depth cues
  - B. monocular depth cues
  - C. Gestalt principles of proximity and continuity
  - D. perceptual principles only operate in "real life" situations
62. Through direct experience with animals, we come to anticipate that dogs will bark and that birds will chirp. This best illustrates:
- A. the law of effect.
  - B. spontaneous recovery.
  - C. respondent behavior.
  - D. associative learning.
63. Which of the following options is not a stimulant?
- A. Nicotine
  - B. Cocaine
  - C. Alcohol
  - D. Ecstasy
64. The eye's "blind spot" is related to:
- A. light rays focusing too far in front of the retina.
  - B. light rays focusing too far behind the retina.
  - C. a cluster of cells around the fovea which contains cones, but no rods.
  - D. an area without receptor cells where the optic nerve leaves the eye
65. The fact that most people tend to perceive visual elements that are closer to each other as being a single group illustrates the Gestalt principle of
- A. similarity
  - B. continuity
  - C. proximity
  - D. good form
66. \_\_\_\_\_ theory of dreams suggests that the brain engages in a lot of random neural activity and dreams make sense of this activity.
- A. information processing
  - B. wish fulfillment
  - C. activation-synthesis
  - D. cognitive development
67. Basing the estimated probability of an event on the ease with which instances come to mind is called the
- A. law of small numbers
  - B. representativeness heuristic
  - C. conjunction fallacy
  - D. availability heuristic
68. In which of the following parts of the brain would a lesion most likely result in aphasia?
- A. corpus callosum

- B. amygdala
  - C. hypothalamus
  - D. Wernicke's area
69. What percentage of people score between 70 and 130 on a typical test of intelligence?
- A. fifteen percent
  - B. thirty-four percent
  - C. sixty-eight percent
  - D. ninety-five percent
70. Savant syndrome lends support to which theory of intelligence?
- A. Charles Spearman's "g"
  - B. the theory of brain size
  - C. Howard Gardner's theory of multiple intelligences
  - D. the theory of emotional intelligence
71. Which of the following is not true?
- A. IQ is influenced by birth order
  - B. IQ is correlated with head size
  - C. Intelligence is context-dependent
  - D. Preservatives, dyes, colorings, and artificial flavors lead to a decline in IQ scores
72. A \_\_\_\_\_ is a self-confirming concern that one will be evaluated based on a negative stereotype.
- A. Confirmation bias
  - B. Stereotype threat
  - C. Self-bias
  - D. Belief perseverance
73. Phoneme is the
- a. the smallest linguistic unit that has semantic meaning
  - b. the smallest segmental unit of sound employed to form meaningful contrasts between utterances
  - c. the set of logical and structural rules that govern the composition of sentences, phrases, and words in any given natural language
  - d. the study of meaning, usually in language
74. Noam Chomsky's theory of language acquisition holds that
- A. children repeat words that they hear frequently.
  - B. people have an inborn universal grammar that makes learning of language easy for children
  - C. chimpanzees and other apes can easily learn language.
  - D. some children's vocabularies are extensive, despite the fact that they may live in poor environments.
75. Schizophrenia is the most common cause of \_\_\_\_\_
- A. drugs
  - B. brain diseases
  - C. depressive disorders
  - D. psychosis
76. According to humanistic therapy, personal distress occurs when
- A. unconscious conflicts threaten to rise to the surface of conscious awareness
  - B. a person engages in negative thinking

- C. there is incongruence between a person's self-concept and reality
  - D. a person is lacking in self-control
77. The thermometer in Mary's room always overstates actual room temperature by exactly 10 degrees. The thermometer has \_\_\_\_\_ reliability and \_\_\_\_\_ validity.
- A. low; high
  - B. high; low
  - C. low; low
  - D. high; high
78. To find a boutique in a big shopping mall, you could systematically search every floor in the mall. This best illustrates problem solving by means of:
- E. the availability heuristic.
  - F. functional fixedness.
  - G. an algorithm.
  - H. the representativeness heuristic.
79. People prefer an economic agenda when high employment rates are provided, but they are against it when the complementary unemployment rates are accentuated. This best illustrates:
- A. framing.
  - B. belief bias.
  - C. representativeness heuristics.
  - D. confirmation bias.
80. The sentence, "The happy the stared weak at exam teacher" would be virtually impossible for an English speaker to understand because the sentence violates the rules of English
- A. psycholinguistics
  - B. syntax
  - C. morphemes
  - D. phonemes
81. \_\_\_\_\_ is defined as a form of communication consisting of simple two-word, noun-verb sentences that adhere to the grammatical standards of the culture's language.
- A. underextension
  - B. telegraphic speech
  - C. semantic slanting
  - D. overregularization
82. \_\_\_\_\_ is a cognitive bias that limits a person to using an object only in the way it is traditionally used.
- A. functional fixedness.
  - B. the availability heuristic.
  - C. belief perseverance.
  - D. the representativeness heuristic.
83. \_\_\_\_\_ is a rule of thumb wherein people judge the probability or frequency of a hypothesis by considering how much the hypothesis resembles available data as opposed to using a Bayesian calculation.
- A. confirmation bias.
  - B. functional fixedness.
  - C. belief perseverance.
  - D. the representativeness heuristic.

84. \_\_\_\_\_ is the redirection of feelings and desires and especially of those unconsciously retained from childhood toward a new object.
- A. active listening.
  - B. free association.
  - C. systematic desensitization.
  - D. transference.
85. Kass has experienced several recurrent attacks of overwhelming anxiety that occurred suddenly and expectedly. She is now concerned about being in the public because she is never sure when or where she will overcome by anxiety. In this case, Kass is displaying symptoms of
- A. bipolar disorder
  - B. obsessive-compulsive disorder
  - C. anxiety disorder
  - D. schizophrenia
86. Dr. Jones is a psychotherapist who is nonjudgmental and who communicates honestly and spontaneously with her clients. She encourages them to talk openly about their concerns but she keeps her interpretation and advice to a minimum. She feels her job is to provide feedback that helps her clients sort out their own feelings. Dr. Jones is most likely
- A. a psychodynamic therapist
  - B. a behavioral therapist
  - C. an unsuccessful therapist
  - D. a client-centered therapist
87. Every summer, Alex used to take a train to visit her cousins. This year, in the three months before his annual visit there were news reports of four train derailments. Even though Alex normally prefers the slow pace of train travel, he decided to fly when she visits his cousin this year, instead of taking the train. Alex's decision has most likely been affected by the
- A. law of small numbers
  - B. representativeness heuristic
  - C. conjunction fallacy
  - D. availability heuristic
88. On a mathematics test, if you remind a group of girls that boys tend to do better on this type of test, it is likely that the girls will do more poorly on the test than they would have had they not been told. This example illustrates
- A. Confirmation bias
  - B. Stereotype threat
  - C. Self-bias
  - D. Belief perseverance
89. The theoretical orientation to personality that emphasizes humans' unique potential for personal growth is
- A. the psychodynamic approach
  - B. the behavioral approach
  - C. the biological approach
  - D. the humanistic approach

90. According to Maslow, human needs can be represented
- using an interlinked chain
  - using a circle
  - in the form of a closed feedback loop
  - in the form a pyramid
91. Maslow suggested that the characteristics of self-actualizing individuals include
- total congruence and a strong superego
  - extroversion, openness and agreeableness
  - spontaneity and a balance between polarities in personality
  - dependence on others and need to belong
92. The behavioral component of an emotional experience could include
- activation of the parasympathetic nervous system
  - a subjective experience that is evaluative
  - an increase in blood pressure and heart rate
  - a smile or a frown
93. One piece of evidence that suggests the expression of some emotions may be innate is the finding that
- the same types of experience make people feel angry, regardless of their cultural background.
  - the same types of experience make people feel happy, regardless of their cultural background.
  - there is universal recognition of some emotional experience
  - people produce facial expressions that are not seen in nonhuman primates

The type of questions:

| Question Number | Type of Question      |
|-----------------|-----------------------|
| 1               | Old-Factual           |
| 2               | New _Factual          |
| 3               | Old-Conceptual        |
| 4               | Conceptual-Factual    |
| 5               | Conceptual-Factual    |
| 6               | New -Factual          |
| 7               | Old-Factual           |
| 8               | New-Conceptual        |
| 9               | Old-Conceptual        |
| 10              | Old-Conceptual        |
| 11              | Conceptual-Conceptual |
| 12              | New-Conceptual        |
| 13              | Conceptual-Conceptual |
| 14              | Old-Factual           |
| 15              | New _Factual          |
| 16              | New -Factual          |

|    |                       |
|----|-----------------------|
| 17 | Old-Factual           |
| 18 | Conceptual-Conceptual |
| 19 | New -Factual          |
| 20 | Factual-Conceptual    |
| 21 | Old-Conceptual        |
| 22 | Factual-Conceptual    |
| 23 | New -Factual          |
| 24 | Conceptual-Conceptual |
| 25 | Factual-Conceptual    |
| 26 | New-Conceptual        |
| 27 | Conceptual-Factual    |
| 28 | Factual-Conceptual    |
| 29 | Old-Conceptual        |
| 30 | Factual-Factual       |
| 31 | Old-Factual           |
| 32 | Conceptual-Factual    |
| 33 | Factual-Factual       |
| 34 | New -Factual          |
| 35 | Old-Factual           |
| 36 | Old-Factual           |
| 37 | Factual-Conceptual    |
| 38 | Conceptual-Conceptual |
| 39 | Conceptual-Factual    |
| 40 | Old-Factual           |
| 41 | Old-Conceptual        |
| 42 | New -Factual          |
| 43 | New-Conceptual        |
| 44 | Factual-Conceptual    |
| 45 | Conceptual-Conceptual |
| 46 | New-Conceptual        |
| 47 | Conceptual-Conceptual |
| 48 | Old-Conceptual        |
| 49 | New-Conceptual        |
| 50 | Factual-Conceptual    |
| 51 | Conceptual-Conceptual |
| 52 | Old-Conceptual        |
| 53 | Factual-Conceptual    |
| 54 | Conceptual-Conceptual |
| 55 | Old-Conceptual        |
| 56 | New-Conceptual        |
| 57 | Conceptual-Conceptual |
| 58 | Conceptual-Conceptual |
| 59 | Conceptual-Conceptual |

|    |                       |
|----|-----------------------|
| 60 | New-Conceptual        |
| 61 | New-Conceptual        |
| 62 | Factual-Conceptual    |
| 63 | Factual-Factual       |
| 64 | Conceptual-Factual    |
| 65 | Conceptual-Factual    |
| 66 | New-Factual           |
| 67 | Conceptual-Factual    |
| 68 | Factual-Factual       |
| 69 | Conceptual-Factual    |
| 70 | Factual-Factual       |
| 71 | new -Factual          |
| 72 | Conceptual-Factual    |
| 73 | Factual-Factual       |
| 74 | Factual-Factual       |
| 75 | Old-Factual           |
| 76 | new -Factual          |
| 77 | New-Conceptual        |
| 78 | Old-Conceptual        |
| 79 | Factual-Conceptual    |
| 80 | Factual-Conceptual    |
| 82 | new -Factual          |
| 83 | Old-Factual           |
| 84 | Conceptual-Factual    |
| 85 | Factual-Factual       |
| 86 | Conceptual-Conceptual |
| 87 | New-Conceptual        |
| 88 | Factual-Conceptual    |
| 89 | New-Conceptual        |
| 90 | Factual-Factual       |
| 91 | Factual-Factual       |
| 92 | Factual-Factual       |
| 93 | Factual-Factual       |



## Appendix G

### Articles Used in Experiment 4

#### A COURT THAT SENTENCES PSYCHOLOGICAL CARE RATHER THAN JAIL TIME

Two years ago, "Mary," a woman diagnosed with a paranoid schizophrenia and accused of loitering, would have waited in jail for 21 days to have her case heard. Then, upon her conviction, she would have served up to six months in a Broward County, Fla., jail.

But thanks to the county's mental health court, instead of serving jail time, she's under court order to receive the mental health care she needs, which in her case includes medication for her schizophrenia and counseling for substance abuse.

Mary is one of 1,345 cases helped so far by the mental health court, the first U.S. court to provide treatment to mentally ill defendants who are arrested for nonviolent misdemeanors, such as loitering or creating a public nuisance. The voluntary program offers a service that is often missing in the judicial system: diagnosis of mental illness and follow-up treatment so that mentally ill defendants stop bouncing from homeless shelters to jail to hospitals and back again.

And the program is noteworthy for another aspect. Those who conduct the screenings and determine which cases should be referred to mental health court are doctoral students from the Center for Psychological Studies at Nova Southeastern University. Currently the only graduate psychology program to offer a court-based practicum, Nova requires students to spend 10 to 20 hours at the courthouse each week in addition to their other program requirements, such as coursework in forensic psychology and family and criminal law.

Before Nova became involved with the mental health court, no one screened defendants for mental illnesses.

"The students are really an essential operational component of the court," says County Court Judge Ginger Lerner-Wren, who presides over Broward County's mental health court.

The court's success has led a handful of other jurisdictions to establish mental health courts, including King County, Wash., Cook County, Ill., and San Bernardino County, Calif. (see related article, page 59).

Meanwhile, Congress is considering legislation that would provide funding to establish additional mental health courts.

Experts say mental health courts are becoming increasingly important as a growing number of mentally ill individuals are incarcerated in jails and prisons instead of being given mental health treatment. Recent statistics show that:

- Three out of four mentally ill inmates have been sentenced to time in prison, jail or probation at least once prior to their sentence, according to a July 1999 report by the U.S. Bureau of Justice Statistics.
- The same report finds that 16 percent of all inmates in state prisons and local jails suffer from mental illness.
- From 25 percent to 40 percent of America's mentally ill will have contact with the criminal justice system, estimates the National Alliance for the Mentally Ill.

Psychology and law enforcement experts blame this mounting social problem on the drop in mental health services available to people with mental illness after states shuttered their mental health hospitals in the late 1960s and early 1970s. At that time, community treatment centers were expected to fill in the gaps but many couldn't because they lacked the resources to provide services. In recent years, funding for outpatient services has gotten even tighter and people with chronic mental illnesses are more apt to fall through the cracks and end up in jail for petty crimes.

The need to address this problem is why programs like Nova's are seen as essential.

"Forensic psychology has become one of the hottest areas of the field," says clinical and forensic psychologist Lenore Walker, EdD, professor, coordinator of the forensic concentration and practicum supervisor at Nova. "We're teaching psychology students how to apply their clinical skills to a legal setting."

Walker has worked with the Broward County Public Defender's Office for seven years and helped to bring Nova and the Broward County courts together.

Since it began two years ago, 44 students have participated in the forensic program. This year, eight students are participating in the year-long practicum.

Practicum students begin their day early, arriving at the jail at 8 a.m. to screen people who were arrested overnight. The experience gives them an up-close and personal view of the court system and defendants.

"We look at the reasons they're arrested," says Allyson Ruha, a third-year Nova student who is currently in the practicum. "Repeated arrests for disorderly conduct, trespassing and prostitution are often linked to mental health problems."

The atmosphere at the jail is initially unsettling, says Ruha, as the defendants haven't showered and those who have been arrested on battery charges are often bruised and bloody.

Before their arraignments, Ruha and other students ask the defendants about prior mental health treatment, whether they've abused drugs or alcohol and if they've had any head injuries. If they identify a defendant with a mental health problem, they alert the public defender and recommend that the defendant be referred to mental health court. If defendants agree to participate in mental health court, students conduct more extensive assessments, performing psychosocial evaluations and sometimes testifying before the judge as to why defendants should receive treatment rather than jail time.

In such cases, the judge usually orders defendants to participate in a treatment program for at least six months. They may be referred to the local mental health center, an outpatient program for substance abuse or a residential treatment program, depending on their needs.

One place defendants may be referred to is OPTIONS, an outpatient program staffed by Nova students. Since January, students have provided individual and group therapy for the 20 women at a treatment center next to the courthouse. OPTIONS can treat up to 40 women with serious mental illness and substance-abuse-related disorders, particularly those who suffer from severe emotional or physical abuse, major depression or post-traumatic stress disorder.

OPTIONS began this year and is funded by a \$226,000 grant from the Bureau of Justice Assistance of the U.S. Department of Justice. The program focuses on empowering participants through cognitive-behavioral therapy, individual and group psychotherapy, medication, skills training and self-care activities. For instance, they're encouraged to practice yoga and meditation, learn computer skills and participate in art and group therapy.

Several members of Congress are hoping to replicate Broward County's mental health court across the country.

Rep. Ted Strickland, PhD (DOhio), and Sen. Mike DeWine (ROhio), introduced legislation to create additional mental health courts. Strickland's bill would provide \$2 million annually over the next five years to create 25 mental health courts across the nation while DeWine's bill would provide \$50 million over the next five years to create 105 mental health courts. Hearings on the proposals are expected to begin this summer.

Both Strickland, a psychologist, and DeWine, a former prosecutor who managed Ohio's prison system when he was the state's lieutenant governor, know firsthand the problems people with mental illness face in jail.

"No one benefits from our current practice of incarcerating nonviolent petty offenders who are in serious need of mental health treatment," says Strickland. "The courts, jail and prisons have become stifled with cases of individuals who are likely to recidivate unless they receive treatment.

Correctional officers are expected to fill the role of mental health professionals when the mentally ill are sent to jail because many of these inmates don't understand why they are there."

### FIGURING OUT PHOBIA

More than 10 million adults in the United States suffer from some sort of phobia, according to the National Institute of Mental Health. These exaggerated fears--whether of spiders, needles, snakes, heights, social situations or even public spaces --can become so all-consuming that they interfere with daily life.

The good news is that over the past several decades, psychologists and other researchers have developed some effective behavioral and pharmacological treatments for phobia, as well as technological interventions.

Now researchers are taking the next step, says psychologist and phobia researcher Arne Öhman, PhD, of the clinical neuroscience department at the Karolinska Institute in Sweden. They are using neuroimaging techniques like positron-emission tomography (PET) and functional magnetic resonance imaging (fMRI) to understand the brain circuitry that underlies phobia and what happens in the brain during treatment.

They're finding that the amygdala--a small, almond-shaped structure in the middle of the brain's temporal lobes--is a key player, and that malfunctions of the amygdala and associated brain structures may give rise to many phobias. Still, researchers have yet to work out the details of how this happens.

"As soon as we know more about what is happening in the brain, then we can fine-tune treatment," Öhman says.

All phobias are anxiety disorders, lumped in the same class as post-traumatic stress disorder and panic disorder, among others. And anxiety disorders are, fundamentally, based on fear.

"What we know about the neurocircuitry and brain basis of fear originally comes from animal research," says psychiatrist Scott Rauch, MD, of Harvard Medical School. Indeed, more than 30 years of research has examined the neurological underpinnings of fear in laboratory rats.

The workhorse paradigm has been the fear conditioning/fear extinction model, Rauch explains. In this model, researchers condition rats to fear a neutral stimulus, like a particular tone, by pairing it with something aversive, like an electric shock. Then, later, the researchers can "extinguish" this fear by repeatedly playing the tone without the accompanying shock. The researchers can use electrodes to record electrophysiological activity in the rats' brains during the fear conditioning or extinction process.

"Using this paradigm, in the past 25 years we've been able to pinpoint pretty precisely where to look for fear in the brain," says New York University psychologist Joseph LeDoux, PhD, a pioneer of this type of research.

What they've pinpointed is the amygdala. LeDoux and others have found that there is a double pathway leading to and from the amygdala. One path leads directly from a frightening sensory stimulus--like the sight of a snake or the sound of a loud crash--to the amygdala in just a few thousandths of a second. A second, slower pathway travels first to the higher cortex before reaching the amygdala.

"The shorter pathway is fast but imprecise," LeDoux explains. "If a bomb goes off, you might not quickly be able to evaluate any of the perceptual qualities of the sound, but the intensity is

enough to trigger the amygdala. If you knew a lot about bombs, then through the cortex pathway you could evaluate the danger, but it will take longer."

The fast pathway, then, is the brain's early warning system, explains LeDoux, and leads to physical manifestations of fear like a racing heart and sweaty palms. The second pathway can override the first, and either lead to conscious feelings of fear or no fear. Studies like these have led researchers to believe that phobias and other anxiety disorders are caused by some type of dysfunction in the amygdala and related brain areas.

The detail and scale of what researchers have learned from animal experiments is extraordinary, according to Rauch. "But the disadvantage is that you have to extrapolate from what you've learned to humans, and particularly to humans with anxiety disorders," he says.

So about a decade ago, researchers began to try to examine the analogous processes in people, using brain-imaging technology such as PET and fMRI.

What they've found has already led to a greater understanding of many anxiety disorders, particularly obsessive-compulsive disorder and post-traumatic stress disorder.

Fewer studies have focused on phobias, Rauch says: "The data there are a little less developed, and the results less cohesive." The first studies, from the early and mid 1990s, were symptom-provocation studies: Researchers would show, say, a snake-phobic person a snake or a picture of a snake, and then use PET scans to examine the brain's reaction.

"Heuristically, it was appealing to believe that these phobic disorders would be related to abnormalities in the fast-track through the amygdala," Rauch says. But in fact the earliest studies--like a 1995 study by Rauch in the *Archives of General Psychiatry* (Vol. 52, No. 1, pages 20-28)--didn't find any evidence of amygdala activation, although some cortical areas that communicate with the amygdala were active.

As measurement and experimental techniques have developed over the past decade, though, the findings have developed as well. For example, fMRI works more quickly than do PET scans, so researchers can examine the brain's reaction to stimuli in a narrower time scale, Rauch explains. In a 2003 study from *Neuroscience Letters* (Vol. 348, No. 1, pages 29-32), for example, psychologist Wolfgang Miltner, PhD, and his colleagues at Friedrich Schiller University in Germany used fMRI to examine spider phobics as they viewed pictures of spiders,

snakes and mushrooms. This time the researchers found that the amygdala was more active in the spider phobics than in control participants.

Other researchers have found that "masking" the phobia stimulus, so that participants see it but are not consciously aware of it, produces interesting results. In a 2004 study in *Emotion* (Vol. 4, No. 4, pages 340-353), Öhman and his colleagues flashed 16 snake and spider phobics with pictures of a snake and a spider, each followed by a neutral picture. The presentation was so fast that the participants were not consciously aware that they had seen the snake or spider. Next, the researchers waited long enough for the participants to consciously register the feared stimuli before presenting the neutral ones.

The researchers found that when the timing did not allow conscious awareness, the amygdala responded to both the phobic and fear-relevant stimuli (fear-relevant stimuli were snake pictures for spider phobics, and vice versa). But when the timing did allow awareness, the amygdala responded only to the phobic stimuli. This suggests, Öhman says, that the amygdala responds immediately to anything that might be threatening, but that with more time to process other areas of the brain suppress the amygdala's initial response.

Finally, some researchers have begun to look particularly at what happens in the brain during and after phobia treatment. Psychologists Tomas Furmark, PhD, Mats Fredrikson, PhD, and their colleagues at Uppsala University in Sweden used PET scans to examine the brain activity of 18 people with social phobia as the people spoke in front of a group. Then, one-third of the participants received nine weeks of cognitive-behavioral therapy, one-third received the selective serotonin reuptake inhibitor Citalopram and one-third received no treatment. The researchers tested the patients again, using the same public speaking task, at nine weeks and again after one year. They found that the activation in the amygdala and related cortical areas at nine weeks could predict which people's symptoms would improve after one year.

Though all of these findings are shaping researchers' understanding of the parts of the brain that give rise to phobia, the picture is far from complete.

"This is a critical area of research for the future," says Rauch.

#### THE MEMORY-TRANSFER EPISODE

It's March 1960, and James V. McConnell, an assistant professor of psychology at the University of Michigan, is convinced that planarians — common flatworms — hold the key to unraveling the mystery of memory. He has decided to condition them to scrunch when a bright light is flashed. Then, he plans to chop them into pieces, feed them to their cannibalistic brothers and see whether the learned behavior is transferred from the trained victim to the naïve recipient. His eventual goal is to demonstrate that the engram — the physical representation of memory — is encoded in the structure of unique forms of RNA much as inherited traits are encoded in one's DNA.

The story of “McCannibal and his Mau Mau” hypothesis has become part of the folklore of psychology. Often used in textbooks as a humorous hook to grab students' attention in chapters devoted to learning and memory, two things are typically included: references to “memory pills” or “professor burgers” and the alleged fact that no one was ever able to truly replicate the findings. Those who did report positive results, the story goes, were poor scientists who either conducted sloppy experiments that lacked proper controls or simply deceived themselves.

It started innocently enough. In 1953, McConnell, a graduate student at the University of Texas, collaborated with Robert Thompson to show that planarians could be classically conditioned.

Thompson received his degree and went to Louisiana State University to work with rats, while McConnell, upon his arrival at Michigan, stuck with worms. He knew that by cutting a planarian across the middle into head and tail sections, each part would regenerate its missing half. But, he wondered, if you conditioned a planarian, which half of the bisected beast would retain the conditioned response? Working with two students in the newly formed Planarian Research Group, McConnell found, to his astonishment and delight, that the regenerated tails showed as much retention — and in some cases more — than the regenerated heads.

After nearly a year of McConnell's wrangling with referees, the paper appeared in the *Journal of Comparative and Physiological Psychology*. In his next experiments, McConnell and the PRG showed that each regenerated part of trained worms cut in several pieces retained the initial training and, more important, a planarian that, after several regenerations, contained none of the structure of the originally trained animal also retained the memory.



These results led McConnell to think more seriously about the chemical nature of memory. To test this notion, he needed to find a way to transfer the putative molecules from a trained to an untrained animal. But how? They tried to graft the head of a trained worm onto the tail of a naïve worm — but the head kept falling off.

Next, they tried grinding up trained worms and injecting them into naïve recipients, but that didn't work, either. The hypodermic needles were too big — getting one inside a flatworm was like trying to impale a prune with a javelin — and if, by chance, the needle was positioned well enough to inject the planarian-puree, it either oozed out or caused the worm to explode.

The answer came in March 1960 when fellow worm runner Jay Boyd Best wrote McConnell about the cannibalistic tendencies of a particular planarian species. McConnell and the PRG ran pilot studies in April and obtained positive results. Each of the next four replications — each run blind to guard against experimenter bias — also produced promising results.

For many, these results were hard to swallow. That McConnell first reported these results in the *Worm Runners Digest*, a journal/magazine he edited that published a mixture of straight science and spoof, did not help his case. Of more importance, the planarian work was not easily replicable. The beasts were difficult to train, and various experimenters — most notably a team working under the patronage of Nobel laureate Mac Calvin at Berkeley — reported their failure to do so.

Theoretical concerns made the work even less palatable. The conventional view held that memory consisted of electrical impulses traveling along specific neural pathways. But the spectacular success of Watson and Crick led some to wonder: If genetic information is stored in nucleic acids and proteins, why not acquired information, as well? Although many neurophysiologists thought this analogy nothing more than a bad pun, a number of molecular biologists, thinking the time ripe to apply their tools and analytic approach to the study of memory processes, began to discuss seriously whether RNA played a pivotal role in memory processes. Expectations ran high, and work proceeded along a number of collateral paths. The smart bet, however, was that if RNA or any other biochemical agents played any role, it was merely to fortify and grease the wheels of neural processing.

McConnell wagered on the long shot. Soon after the cannibalism experiments, he successfully injected naïve worms with RNA taken from those trained to negotiate a maze and

reported that the training had transferred. He interpreted these findings as providing evidence that specific memories are encoded in the nervous system in the form of unique structural variants of RNA.

The cannibalism studies, both startling and vivid in their imagery, and McConnell, never one to shy away from the media, caught the public eye. At a time when scientists remained sequestered in their labs, McConnell appeared with his cannibalistic worms on television (i.e., “The Way Out Men,” “Mr. Wizard” and “The Steve Allen Show”), while articles profiling his work appeared in *Time*, *Newsweek*, *Life*, *Esquire* and *Fortune*. Eminently quotable, McConnell referred to his work as confirming the Mau Mau hypothesis, and the “McCannibal” moniker didn’t bother him one bit. He made grand pronouncements about the future of “memory pills” and “memory injections,” promising more than he and others working in the area could actually deliver.

None of this endeared McConnell to his critics.

Still, McConnell believed that eventually the data would win out, and many eminent psychologists, Donald Hebb, Harry Harlow, Karl Pribram, and Gordon Bower among them, fully supported his efforts, even though they did not share his interpretation of his results. In fact, up until 1965, McConnell was, as he put it, “riding high.” He was invited to share a platform with top-flight molecular biologists and electrophysiologists at conferences at the University of California, Los Angeles, in 1962 and Princeton in 1963. During the period from 1959 through 1964, he received more than \$150,000 (in 1960s dollars) from the Atomic Energy Commission and the National Institute of Mental Health designated specifically for the planarian work. He was offered a fellowship to spend a year at the newly created Center for Advanced Study in the Behavioral Sciences in Palo Alto, Calif., in 1960, and he received a prestigious five-year Research Career Development Award from the National Institutes of Health, 1963–68. He received accelerated promotion to full professor at Michigan in 1963.

Everything changed when, in late 1965, four independent labs reported successful memory-transfer experiments using rats (and in one case, cross-species transfer between rats and mice). Two of these reports appeared in the high-impact journals *Science* and *Nature*.

No one could argue that rats cannot learn. Within a few months, more than 50 labs, including teams at Berkeley, Harvard, MIT and Yale, conducted transfer experiments. McConnell, after failed attempts using salamanders and mynah birds, also turned to rats.

### THE BRAIN IN CONTROL

The ability to control our thoughts and behavior is a fundamental human faculty. However, researchers have yet to pinpoint how the soft tissues and electronic currents that make up the brain dictate our thoughts and influence our actions. Indeed, even neuroscientists still resort to metaphysical theories to explain the connection.

"This is a fundamental 'holy grail' problem in neuroscience and psychology," says cognitive neuroscientist Todd Braver, PhD, "We feel that we are in control of our own behavior, but yet when we try to understand how that control emerges out of the neural components of the brain, the physical tissue, we end up reverting to the idea of a homunculus-that there's this little man in the head who's making the key decisions and doing the most important control operations."

Braver, associate professor and co-director of the cognitive control and psychopathology laboratory at Washington University in St. Louis, has devoted his career to banishing the notion of a homunculus in psychological and neuroscience theories. He aims to discover the neural mechanisms behind cognitive control-the ability to form, maintain and realize internal goals. Braver uses a combination of brain imaging, computational modeling and behavioral studies to investigate how people self-regulate their thoughts and behaviors across a range of tasks involving memory, attention and decision-making.

In recognition of his accumulated research accomplishments, as well as his application of his findings to clinical populations such as aging adults or people with schizophrenia or Alzheimer's disease, in August, the American Psychological Foundation (APF) awarded him the \$25,000 APF F.J. McGuigan Young Investigator Research Prize. APF gives the biennial prize to a psychologist less than nine years postdoctorate who conducts psychophysiological research.

Braver credits his interest in psychology in part to his family heritage: His father, Sanford Braver, PhD, is a social psychology professor at Arizona State University; his mother was a clinical psychologist and social worker; and his grandfather was a psychiatrist. Even with such a strong legacy, though, Braver wasn't initially attracted to psychology as an undergraduate.

"I was somewhat resistant to the idea of being in psychology," he says. "I wanted to be in hard science because I naively thought that psychology was too mushy. I thought I wanted to be a physicist."

Braver began school at the University of California, San Diego (UCSD), and after realizing he wasn't cut out to be a physicist, he was drawn to the field of cognitive science by his interest in quantitative precision and the use of physical principles to understand the mind. UCSD was, he says, an institution doing cutting-edge work in cognitive neuroscience, and biologically based computational modeling of cognition.

Braver went on to receive both his MS and PhD degrees in cognitive neuroscience from Carnegie Mellon University, where he also studied at the Center for the Neural Basis of Cognition (CNBC) before beginning his current post at Washington University in St. Louis in 1998.

Braver was attracted to the CNBC, a joint project between the University of Pittsburgh and Carnegie Mellon University, because of its interdisciplinary approach to cognitive neuroscience. The program combines psychology, neuroscience, computer science and mathematics.

"The CNBC was a wonderful place to learn how to link the mind and the brain from many different perspectives," Braver says, noting that this approach influences his work to this day.

Much of Braver's current research focuses on his new theory of cognitive control strategy, which he calls the dual mechanisms of control (DMC) model. Braver has found that cognitive control is either proactive or reactive. For example, a person can proactively control a plan to stop at the grocery store during a drive home by keeping in mind the goal of getting to the store before approaching a turnoff that leads to it. Keeping the goal actively in mind could make driving behavior more effective by ensuring that the car is in the correct turning lane before the intersection is reached. However, even if the person forgets the grocery store stop, reactive control can still kick in when the intersection is reached if, for example, the left turn light triggers a reminder of the goal.

Both forms of cognitive control have their benefits, says Braver. Proactive control is generally more effective, but also demands more energy and is more vulnerable to interruptions.

Reactive control, though, is more susceptible to interference effects, but is also less demanding than proactive control.

People use both proactive or reactive control, adds Braver, but may have a tendency to favor one form or another depending on the specific situation, or through more trait-like biases.

"We've done a number of experiments that show you can manipulate the tendency of one of these mechanisms or another to be used, and they are not only related to properties of the task, but may also be impacted by stable individual differences that people have," says Braver. "We've been looking at cognitive-related individual differences, as well as personality-related individual differences. Both of these factors may have an important influence on the type of control strategies people use in cognitive situations."

In addition to teasing out different control strategies in healthy populations, Braver is also interested in the clinical applications of his findings in populations such as people with schizophrenia or Alzheimer's disease. He runs a joint Washington University lab with his wife, Deanna Barch, PhD, associate psychology professor and assistant psychiatry professor, who studies the neurobiological mechanisms that contribute to language and cognitive deficits in people with schizophrenia, and other clinical populations.

"Through my interactions with Deanna and her group, I continually get informed and influenced by issues that arise in clinical populations, such as schizophrenia. These issues have enormous implications for our understanding of the normal functioning brain and mind," says Braver. "It's important not only to do basic research, but also to apply what we learn in clinical domains. For example, I'm excited by the prospect that we can use the DMC model to do better cognitive training in groups that have well-known problems with cognitive control, such as older adults and patients with schizophrenia."

Braver says his future plans will perpetuate the multidimensional approach to understanding the brain that he learned at UCSD and CNBC: In collaboration with colleagues domestically and internationally, he aims not only to elaborate on his DMC model and his cognitive training research in older adults and people with schizophrenia, but will also focus more closely on decision-making tasks as another way of understanding cognitive control.

"There is so much good work going on outside of America," notes Braver. "One of the things the McGuigan money is allowing me to do is visit other labs more regularly and start to form collaborations with international colleagues."

### STEMMING SOCIAL PHOBIA

As a graduate student at Florida State University, Richard Heimberg, PhD, was fascinated by the effects of people's intense social anxiety on their relationships. However, in the 1970s, no name existed for this type of anxiety, and Heimberg says many people wrote it off as shyness or a personality trait.

After all, many people get a little anxious when delivering a speech to a crowd. But it's the more severe cases that cause Heimberg concern--when those fears of being judged by others become so persistent and intense that they extend to almost all social situations, from informal conversations to eating in public. People diagnosed with the condition, social phobia--also known as social anxiety disorder--may avoid many social situations out of fear that others will notice something unusual about them, like their shaking hands or blushing, and that their actions will embarrass or humiliate them.

To help them overcome the condition, Heimberg, a Temple University psychology professor, has made studying the origins of and treatments for social phobia his life's work.

In 1983, he became the first researcher to receive National Institute of Mental Health (NIMH) funding to study psychosocial treatments for social phobia after the term first appeared in the third edition of the Diagnostic Statistical Manual of Mental Disorders in 1980 (DSM-III).

"Dr. Heimberg has made huge contributions to social phobia research, developing a cognitive-behavioral treatment for social phobia and carrying out numerous randomized controlled trials that have demonstrated its effectiveness," says psychologist Jacqueline Persons, PhD, a former president of the Association for Advancement of Behavior Therapy (AABT) who served with Heimberg on the AABT board. "He has made important contributions to alleviate a great deal of suffering."

Since Heimberg secured the first NIMH social phobia treatment research grant, such funding has been readily available because the condition is quite common: It's the third most

prevalent mental disorder behind depression and alcoholism. About 5.3 million American adults have social phobia, which usually begins in childhood or adolescence, according to NIMH.

For many social phobics, preoccupation with what others think may interfere in their job, school, relationships or other social activities.

"Everyday interactions can become very problematic for people with social anxiety disorder," says Heimberg who, as director of Temple University's Adult Anxiety Clinic, helps people change their thought processes in such interactions using cognitive-behavioral therapy (CBT) and medication. The treatment also encourages clients to expose themselves gradually to feared events.

Since 1983, Heimberg has conducted a series of NIMH-funded studies on social phobia. Most recently, he received a \$1.2 million, four-year NIMH grant--which he's in the second year of--to investigate whether the addition of CBT to medication treatments can help prevent patients' relapse.

Through his 20-plus years of research, Heimberg has found that nearly everyone fears social situations to some degree.

"Some people just think they are shy--that it's a personality trait--and that's just the way they are," says Heimberg. "But...if a person starts fearing many social situations, [and as a result] lives alone or drops out of school, that's not shyness--that's an impairment."

What's more, notes Heimberg, social phobia is generally more debilitating than phobias focused on singular circumstances, such as a fear of thunderstorms or animals. "If you are afraid of interacting with people, that can mess you up wherever you turn," he says. "It can have very broad mental health implications."

Heimberg also notes two subtypes of social phobia. For people with a "generalized" type, the social anxiety ranges across a broad number of social interactions; for those with a "specific" type, the anxiety involves only one or a few social encounters, such as public speaking or eating in public.

The trouble is, despite the proven efficacy of treatment, many social phobics shy away from it, according to Heimberg's research. For example, he found that 92 percent of people who were accessing information on social phobia on an anxiety clinic Web site met criteria for social

anxiety disorder. Yet, only about 36 percent of the respondents reported receiving psychotherapy; 35 percent reported taking medication for social anxiety disorder, according to a study by Heimberg and psychologists Brigitte Erwin, PhD, Cynthia Turk, PhD, David Fresco, PhD, and Donald Hantula, PhD, in the 2004 issue of the *Journal of Anxiety Disorders* (Vol. 18, No. 5, pages 629-646).

But, with CBT treatment or antidepressant medication, about 80 percent of social phobics can alleviate their symptoms, Heimberg says. What's more, clients who receive CBT treatment remain improved five years later, whereas clients who receive only medication treatment are more likely to relapse than clients receiving CBT, according to Heimberg's 1998 study in the *Archives of General Psychiatry* (Vol. 55, No. 12, pages 1,133-1,141).

However, Heimberg suspects a combination of CBT and medication may prove most effective in preventing relapse--something he hopes to prove in his latest NIMH study with Michael Liebowitz, MD, of the New York State Psychiatric Institute, to be completed by 2007.

Heimberg is providing participants with a 28-week treatment program that includes medication--in this case, the antidepressant Paxil--and some patients then receive 16 sessions of CBT. Each session helps clients to evaluate their thought processes more critically, such as by filtering out automatic thoughts that others judge them negatively in social interactions. The client and therapist also role-play social interactions to provide the client with confidence they can take into real-world situations. In Heimberg's CBT, the therapist gradually exposes clients to their feared social situations in real life, perhaps assigning a client to initiate a conversation with a person they don't know, ask someone out on a date or go on a job interview.

"They start going into social situations that have made them tense a thousand times before, but the trick now is that they are doing it...with coping skills that will help them turn defeat into victory," Heimberg says.

#### TOOL TALK

Sally Boysen, PhD, remembers the day in 1995 when renowned animal researcher Gordon Gallup, PhD, was taken aback by seeing one of her chimpanzees eating yogurt with a spoon. Here was a father of animal self-awareness research, who knew that chimps were master tool



users, spellbound by the chimp's simple act. Knowing animals are capable of tool use is one thing, but seeing it is quite another, Boysen says.

"He acted like she'd just recited the Gettysburg Address," says Boysen, a cognitive psychologist at The Ohio State University.

Since the 1950s, a steady stream of research has shown that tool use isn't unique to humans. Chimps use and modify tools to dig out termites from holes and smash tough nuts. Tool use has also been observed in captive baboons, gorillas and bonobos, wild and captive orangutans, and capuchin and macaque monkeys. Even nonprimates, including elephants, crows and woodpecker finches, create and use tools.

Still, it's been unclear what these species have in common that allows them to manipulate objects in their environment to their advantage. But new research by psychologists is changing that. Psychologists have found similarities in brain structure and a common cognitive ability to represent abstract concepts, as well as an interesting evolutionary parallel between language and tool use in humans. Far from a mystifying and unique skill, these findings reveal tool use to be a quite common — albeit incredibly adaptable — strategy a number of animals have evolved to get the most out of their environments.

Chaminade, at the Mediterranean Institute for Cognitive Neuroscience in Marseilles, France, and Stout, at Emory University in Atlanta, published a meta-analysis last year in the *Cambridge Archaeological Journal* (Vol. 19. No. 1) arguing that in humans, language has developed hand-in-hand with the ability to use and create increasingly complex tools. Both skills require similar cognitive architecture and break down into structured sequences of behavior, they say.

There's also overlap in at least one brain region, they add. The premotor cortex plays a big role in perceiving and making sense of language sounds, as well as telling the hands what to do when grasping, flexing and rotating. In a way, that overlap makes sense, they say. Both grasping and sound processing take discrete gestures — individual muscle movements of the fingers, wrist and arm; the individual sounds produced by the tongue, lips and palate — and merge them into something cohesive and meaningful, they say.

So would making a stone tool have activated the premotor cortexes in our early ancestors? Chaminade and Stout tried to find that out through a modern-day recreation of a Stone Age task. They asked expert stone tool makers to create a stone hand axe by chipping away one stone with another during a PET scan. As they suspected, the researchers found a high degree of activation in the tool-makers' premotor cortexes.

As tools became an important aspect of human societies, evolution would have selected for expanded premotor cortexes and other brain regions responsible for tool use and creation. That, in turn, may have expanded the areas necessary for more complex language. And as language became an important trait in its own right, it too would be affected by evolutionary force, further accelerating expansion of these brain regions. Tool use reinforces language and vice versa, and evolution selects for both traits.

Even though chimps can't program a computer or build skyscrapers, new research finds they have increasingly impressive tool-using abilities. In one such study, Boysen has confirmed a behavior that chimp researchers have long suspected but never proven in a lab: Chimps not only know how to use tools, they can modify existing tools to suit different situations as they arise.

Boysen and her colleagues ran an experiment with nine chimps — five adults, two juveniles and two infants. The chimps were presented with one of two contraptions containing food, as well as pieces of PVC piping that they could manipulate into various tools. The first was a two-level box with a shelf separating the levels. By inserting a stick through a hole in the side of the box, the chimp could push a food reward off the shelf and down into the lower level, where he or she could pick it up and eat it. The second contraption was a flat plank, just out of reach, with a food reward on it and a dowel sticking up from it.

The PVC tool in its deconstructed state was simply a straight pipe with a T-joint at the ends, narrow enough to fit into the hole of the two-level box. By attaching dowels at the T-joint, the chimps could form a hook to catch the dowel and pull close the wooden plank. The researchers presented the chimps with one of the contraptions and the tool in either its constructed or deconstructed state, observed whether the chimps recognized the tool needed to be changed or kept the same, then repeated the process.

The adults and juveniles aced the test, adjusting the tool to its correct configuration nearly every time. The infants, though, were much less accurate, performing at or below chance when initially given a tool configuration that didn't fit the contraption.

The fact that the adults and juveniles so effortlessly constructed and deconstructed the proper tool for the job seems to indicate a causal understanding of what a tool is and how it can be used, Boysen says. In fact, the chimps with their PVC pipes appear to be using the same cognitive resources you'd use to put together a bookshelf.

And those cognitive resources are likely an evolutionary extension of ordinary behavior, Boysen says.

And, like in human infants, chimp infants need time and experience for their brains to develop enough to understand these mental representations. Both study their parents' behaviors for years before they learn the ropes.

That explanation accounts for our evolutionary cousins, but it doesn't explain why birds can use tools. The hyacinth macaw, a parrot native to central and eastern South America, for example, uses wooden wedges to pop open nuts, says Douglas Wylie, PhD, a cognitive psychologist at the University of Alberta in Edmonton. New Caledonian crows fashion spears out of twigs, which they hold in their beaks to stab for grubs in trees.

So why can humans, chimps and certain birds use tools, while so many other species can't? Wylie says part of the answer may lie in the cerebellum, home to motor control and motor learning, two cognitive skills necessary for tool use. Numerous studies over the last decade have looked to that brain region as the possible deciding factor in tool use, but they have yet to find any correlation between cerebellar volume and species that use tools. But it's not just volume that counts, Wylie says. Brains also vary in the degree to which their cortexes fold. The more folding, the more connections among neurons, as well as increased surface area for the same amount of space.

To Wylie, that suggests a heavily folded cerebellar cortex "may be a common means of achieving proficient tool use," he says.

## Appendix H

### The Questions Used In Experiment 4

#### Multiple Choice Questions

#### **The Memory-Transfer Episode**

1. What is an engram?
  - a. The structure of unique forms of RNA
  - b. Inherited traits that are encoded in one's DNA.
  - c. A type of memory pill.
  - d. The physical representation of memory
2. What is "McCannibal and his Mau Mau Mau" hypothesis?
  - a. Learned behavior can be transferred to new recipients.
  - b. Rats can learn and transfer what they have learned.
  - c. The procedure of removing the head of a trained worm and grafting it onto the tail of a naïve worm.
  - d. Cannibalistic tendencies of a particular planarian species.
3. What is the conventional view of memory?
  - a. Genetic information is stored in nucleic acids and proteins.
  - b. RNA plays an important role in memory processes.
  - c. Memory is a result of electrical impulses traveling along neural pathways.
  - d. Memory is the result of generation of new neurons.
4. What is the most important theoretical assumption of cannibalism experiments?
  - a. Genetic information is stored in nucleic acids and proteins,
  - b. RNA plays an important role in memory processes.
  - c. Memory is a result of electrical impulses traveling along neural pathways.
  - d. Memory is the result of generation of new neurons.
5. What might be the future consequences of Mau Mau hypothesis?
  - a. Development of memory injections and memory pills.
  - b. Storing the information in individual neurons.
  - c. Transfer of information from one individual to the other.
  - d. Development of biochemical agents to improve memory.
6. What is the reason that researchers prefer to use rats in their experiments?
  - a. Because the transfer studies failed with other species.
  - b. Because they can learn faster than other species.
  - c. Because rats' intellectual abilities are greater than the other species.
  - d. Because rats' life span is longer than other laboratory animals.

#### **Figuring out Phobia**

1. What is phobia?
  - a. Exaggerated fear
  - b. Irrational fear
  - c. Unexplainable anxiety

- d. The response to unexpected environmental trigger
2. What are the treatments for phobia?
  - a. Electroconvulsive therapy and hypnosis
  - b. Group therapy and medicine
  - c. Desensitization and behavioral therapies
  - d. Behavioral and pharmacological treatments.
3. What is the brain structure which is associated with phobias?
  - a. Cerebellum
  - b. Temporal lobe
  - c. Amygdala
  - d. Cortical areas
4. Why is it important to know underlying brain mechanism of phobias?
  - a. To understand the symptoms of phobias
  - b. To differentiate between different types of phobias
  - c. To develop a theory to explain phobias
  - d. To develop proper treatment.
5. What is the name of the paradigm to study phobias?
  - a. Fear conditioning/fear extinction paradigm
  - b. Workhorse paradigm
  - c. Masking Paradigm
  - d. Cognitive-Behavior Paradigm
6. What is the function of the fast pathway?
  - a. To make mental calculations about the feared object
  - b. To lead to physical symptoms of fear
  - c. To trigger the proper brain regions
  - d. To initiate the fight or flight response

### Tool Talk

1. Which animals are reported to use tools?
  - a. Crows
  - b. Flies
  - c. Raccoons
  - d. Snakes
2. What are the two ways that chimps use and modify tools?
  - a. To hunt and to fight with an opponent
  - b. To dig out termites and to smash tough nuts
  - c. To reach for food just out of reach and to hunt
  - d. To do extracted foraging and for mating
3. Compared to other species, how did humans get so tool proficient?
  - a. Because humans have bigger frontal lobe.
  - b. Because of evolutionary pressure.
  - c. Because human have the ability to acquire language.
  - d. Because human have cognitive ability to manipulate tools.
4. Which brain region is responsible for tool use?
  - a. The premotor cortex
  - b. The Cerebellum
  - c. The motor cortex

- d. The temporal lobe
- 5. Why can human and chimps and certain animals use tools, while so many other species cannot?
  - a. Because of brain size.
  - b. Because their learning capacity.
  - c. Because the cerebellum is more folded.
  - d. Because the premotor cortex is more folded.
- 6. Why are human infants and chimp infants not capable of tool use?
  - a. They need to engage in observational learning to learn how to use tools.
  - b. Their brains need to develop enough to understand mental representations.
  - c. The neural connections are not established yet.
  - d. They cannot engage in mental extraction.

**A court that sentences psychological care rather than jail time**

- 1. Who conduct/s the screenings of defendants and determine which cases should be referred to mental health court?
  - a. Licensed psychologists
  - b. The judge
  - c. Doctoral students
  - d. Prosecutor
- 2. What is the function of the mental health court?
  - a. Independent of the nature of crime, screening defendants for mental illness.
  - b. Providing psychological help to defendants for their entire life.
  - c. Improving defendants' social functioning by treating the mental problems.
  - d. Differentiating the causes of different nonviolent misdemeanors.
- 3. What is percentage of inmates who are in state prison suffering from mental illness?
  - a. 25%
  - b. 40%
  - c. 16%
  - d. 30%
- 4. What is the reason of arrest which correlates with mental health problems?
  - a. Prostitution
  - b. Murder
  - c. Vandalism
  - d. Aggravated assault
- 5. What is OPTIONS?
  - a. Local mental health center
  - b. An outpatient program
  - c. A graduate program
  - d. A state hospital
- 6. What are the disorders that OPTIONS aims to treat?
  - a. Schizophrenia
  - b. ADHD
  - c. Obsessive-compulsive personality disorder
  - d. Major depression

### The Brain in Control

1. According to the passage, what is the fundamental problem in neuroscience and psychology?
  - a. How does human psyche function?
  - b. How do we contract memory?
  - c. How does human brain operate?
  - d. How do we control our cognitive processes?
2. What does the homunculus represent?
  - a. Little man in the head
  - b. Driving force of human psyche
  - c. The connection between the mind and the brain
  - d. the electrochemical communication between neurons
3. What is the purpose of the CNBC?
  - a. To investigate the neural processes
  - b. To explain mind-body problem
  - c. To solve methodological problems faced by psychologist and neuroscientists
  - d. To find the link between the mind and the brain.
4. What is the most important problem associated with reactive control?
  - a. It is very susceptible to interference effects.
  - b. It is very vulnerable to interruptions.
  - c. It is affected by mood.
  - d. It is not reliable.
5. What is the most important problem associated with proactive control?
  - a. It is very susceptible to interference effects.
  - b. It is very vulnerable to interruptions.
  - c. It is affected by mood.
  - d. It is not reliable.
6. What methods are used to investigate how people self-regulate their thoughts and behavior?
  - a. Medicine, therapy and brain imaging
  - b. Hypnosis, medicine, computational modeling
  - c. Cognitive therapy, brain imaging and behavioral studies
  - d. Brain imaging, computational modeling and behavioral studies

### Stemming social phobia

1. Before the term “social phobia” was coined, which term was used to describe social phobia?
  - a. Anxiety disorder
  - b. Shyness
  - c. Personality disorder
  - d. Nervousness
2. Which aspect of one’s life can be negatively affected by social phobia?
  - a. Eating in a restaurant
  - b. Playing a guitar
  - c. Problem solving
  - d. Attention

3. When did the term “social phobia” appear for the first time?
  - a. DSM-III in 1981
  - b. DSM-II in 1980
  - c. DSM-III in 1980
  - d. DSM-IV in 1980
4. What types of treatments are used to help people with social phobia?
  - a. Clinical interventions
  - b. Electroconvulsive therapy
  - c. Hypnosis and medication
  - d. Cognitive-behavioral therapy and medication
5. What is the generalized type of social phobia?
  - a. The anxiety involves only one or a few social encounters.
  - b. The anxiety is generalized to singular circumstances such as fear of animals.
  - c. The anxiety is extended to all the aspects of one’s life.
  - d. The anxiety ranges across a number of social interactions.
6. What does the proposed 28-week treatment program for social phobia aim to do?
  - a. To treat social phobia
  - b. To erase the symptoms of social phobia
  - c. To prevent relapse
  - d. To teach the patients the coping skills

### **Short Answer Questions**

#### **The Memory-Transfer Episode**

1. What is an engram?
2. What is “McCannibal and his Mau Mau Mau” hypothesis?
3. What is the conventional view of memory?
4. What is the most important theoretical assumption of cannibalism experiments?
5. What might be the future consequences of Mau Mau hypothesis?
6. What is the reason that researchers prefer to use rats in their experiments?

#### **Figuring out Phobia**

1. What is phobia?
2. What are the treatments for phobia?
3. What is the brain structure which is associated with phobias?
4. Why is it important to know underlying brain mechanism of phobias?
5. What is the name of the paradigm to study phobias?
6. What is the function of the fast pathway?

#### **Tool Talk**

1. Which animals are reported to use tools?
2. What are the two ways that chimps use and modify tools?
3. Compared to other species, how did humans get so tool proficient?



4. Which brain region is responsible for tool use?
5. Why can human and chimps and certain animals use tools, while so many other species cannot?
6. Why are human infants and chimp infants not capable of tool use?

#### **A court that sentences psychological care rather than jail time**

1. Who conduct/s the screenings of defendants and determine which cases should be referred to mental health court?
2. What is the function of the mental health court?
3. What is percentage of inmates who are in state prison suffering from mental illness?
4. What is the reason of arrest which correlates with mental health problems?
5. What is OPTIONS?
6. What are the disorders that OPTIONS aims to treat?

#### **The Brain in Control**

1. According to the passage, what is the fundamental problem in neuroscience and psychology?
2. What does the homunculus represent?
3. What is the purpose of the CNBC?
4. What is the most important problem associated with reactive control?
5. What is the most important problem associated with proactive control?
6. What methods are used to investigate how people self-regulate their thoughts and behavior?

#### **Stemming social phobia**

1. Before the term “social phobia” was coined, which term was used to describe social phobia?
2. Which aspect/s of one’s life can be negatively affected by social phobia?
3. When did the term “social phobia” appear for the first time?
4. What types of treatments are used to help people with social phobia?
5. What is the generalized type of social phobia?
6. What does the proposed 28-week treatment program for social phobia aim to do?